CHAPTER 1
Thinking Critically With Psychological Science

Eva-Katalin/Getty Images

Hoping to satisfy their curiosity about people and to relieve their own woes, millions turn to "psychology." They read advice columns aimed at helping people cope with their problems, overcome their addictions, and save their marriages. They watch “celebrity psychics” demonstrate their supposed powers. They attend stop-smoking hypnosis seminars. They play online games, hoping to strengthen their brain. They immerse themselves in self-help books, websites, and lectures that promise to teach the path to love, the road to personal happiness, and the “hacks,” or shortcuts, to success.

Others, intrigued by claims of psychological truth, wonder: How—and how much—does parenting shape children’s personalities and abilities? Are first-born children more driven to achieve? Do dreams have deep meaning? Do we sometimes remember events that never happened? Does psychotherapy heal?

In working with such questions, the science of psychology does more than speculate. To separate uninformed opinions from examined conclusions, psychologists use the *scientific method* to conduct research. Let’s consider how psychology’s researchers do their science.

Research Strategies: How Psychologists Ask and Answer Questions

The Need for Psychological Science

LEARNING OBJECTIVE QUESTION LOQ 1-1

How does our everyday thinking sometimes lead us to a wrong conclusion?

Some people think that psychology merely proves what we already know and then dresses it in jargon: “You get paid for using fancy methods to tell me what my grandmother knew?”

Indeed, Grandma’s common sense is often right. As the baseball great Yogi Berra (1925–2015) once said, “You can observe a lot by watching.” (We also have Berra to thank for other gems, such as “Nobody goes there anymore—it’s too crowded,” and “If the people don’t want to come out to the ballpark, nobody’s gonna stop ’em.”) Because we’re all behavior watchers, it would be surprising if many of psychology’s findings had *not* been foreseen. Many people believe that love breeds happiness, for example, and they are right (we have what Chapter 11 calls a deep “need to belong”).

But sometimes Grandma’s common sense, informed by countless casual observations, is wrong. In later chapters, we will see how research has overturned popular ideas—that familiarity breeds contempt, that dreams predict the future, and that most of us use only 10 percent of our brain. We will also see how research has surprised us with discoveries about how the brain’s chemical messengers control our moods and memories, about other animals’ abilities, and about the effects of stress on our capacity to fight disease.

Other things seem like commonsense truth only because we so often hear them repeated. Mere repetition of statements—whether true or false—makes them easier to process and remember, and thus more true-seeming (Dechêne et al., 2010; Fazio et al., 2015). Easy-to-remember misconceptions ("Vitamin C prevents the common cold") can therefore overwhelm hard truths. This power of familiar, hard-to-erase falsehoods is a lesson well known to political manipulators, and kept in mind by critical thinkers.

> **"All effective propaganda must be limited to a very few points and must harp on these in slogans."** — Adolf Hitler, *Mein Kampf*, 1926

Three common flaws in commonsense thinking—*hindsight bias, overconfidence,* and *perceiving order in random events*—illustrate how, as novelist Madeleine L'Engle (1973) observed, "The naked intellect is an extraordinarily inaccurate instrument."

> **"Those who trust in their own wits are fools."** — Proverbs 28:26

Did We Know It All Along? Hindsight Bias

Consider how easy it is to draw the bull's-eye *after* the arrow strikes. After the stock market drops, people say it was "due for a correction." After the athletic match, we credit the coach if a "gutsy play" wins and criticize the same "stupid play" if it doesn't. After a war or an election, its outcome usually seems obvious. Although history may therefore seem like a series of inevitable events, the actual future is seldom foreseen. No one's diary recorded, "Today the Hundred Years War began."

> **"Life is lived forwards, but understood backwards."** — Philosopher Søren Kierkegaard, 1813–1855

This **hindsight bias** is easy to demonstrate by giving half the members of a group some purported psychological finding and giving the other half an opposite result. Tell the first group, for example: "Psychologists have found that separation weakens romantic attraction. As the saying goes, 'Out of sight, out of mind.'" Ask them to imagine why this might be true. Most people can, and after hearing an explanation, nearly all will then view this true finding as unsurprising.

hindsight bias
the tendency to believe, after learning an outcome, that one would have foreseen it. (Also known as the *I-knew-it-all-along phenomenon.*)

Tell the second group the opposite: “Psychologists have found that separation strengthens romantic attraction. As the saying goes, ‘Absence makes the heart grow fonder.’” People given this untrue result can also easily imagine it, and most will also see it as unsurprising. When opposite findings both seem like common sense, there is a problem.

> **“Anything seems commonplace, once explained.”** — Dr. Watson to Sherlock Holmes

Such errors in people’s recollections and explanations show why we need psychological research. It’s not that common sense is usually wrong. Rather, common sense describes, after the fact, what *has* happened better than it predicts what *will* happen.

More than 800 scholarly papers have shown hindsight bias in people young and old from around the world (Roese & Vohs, 2012). As physicist Niels Bohr reportedly jested, “Prediction is very difficult, especially about the future.”

Everett Collection/Newscom

Hindsight bias When drilling its Deep-water Horizon oil well in 2010, BP employees took shortcuts and ignored warning signs, without intending to harm people, the environment, or their company's reputation. *After* an explosion killed 11 employees and caused the largest ever marine oil spill, the foolishness of those judgments became (in hindsight) obvious.

Overconfidence

We humans tend to think we know more than we do. Asked how sure we are of our answers to factual questions *(Is Boston north or south of Paris?)*, we tend to be more confident than correct.[1] Consider these three anagrams, shown beside their solutions (from Goranson, 1978):

WREAT → WATER
ETRYN → ENTRY
GRABE → BARGE

About how many seconds do you think it would have taken you to unscramble each of these? Did hindsight influence you? Knowing the answers tends to make us overconfident. (Surely the solution would take only 10 seconds or so?) In reality, the average problem solver spends 3 minutes, as you also might, given a similar anagram without the solution: OCHSA.[2]

Fun anagram solutions from Wordsmith (wordsmith.org):
Snooze alarms = Alas! No more z's
Dormitory = dirty room
Slot machines = cash lost in 'em

Are we any better at predicting social behavior? Psychologist Philip Tetlock (1998, 2005) collected more than 27,000 expert predictions of world events, such as the future of South Africa or whether Quebec would separate from Canada. His repeated finding: These predictions, which experts made with 80 percent confidence on average, were right less than 40 percent of the time. It turns out that only about 2 percent of people do an excellent job predicting social behavior. Tetlock (with Gardner, 2016) calls them "superforecasters." Superforecasters avoid overconfidence. Faced with a difficult prediction, a superforecaster "gathers facts, balances clashing arguments, and settles on an answer."

Overconfidence in history: "We don't like their sound. Groups of guitars are on their way out." — Decca Records, in turning down a recording contract with the Beatles in 1962

"Computers in the future may weigh no more than 1.5 tons." — *Popular Mechanics*, 1949

"They couldn't hit an elephant at this distance." — General John Sedgwick just before being killed during a U.S. Civil War battle, 1864

"No woman in my time will be prime minister." — Margaret Thatcher, 1969 (British Prime Minister, 1979–1990)

ASK YOURSELF

Do you have a hard time believing you may be overconfident? Could overconfidence be at work in that self-assessment? How might reading this section about overconfidence help reduce your tendency to be overconfident?

RETRIEVAL PRACTICE

RP-1 Why, after friends start dating, do we often feel that we *knew* they were meant to be together?

ANSWERS IN APPENDIX E

Perceiving Order in Random Events

We're born with an eagerness to make sense of our world. People see a face on the Moon, hear Satanic messages in music played backward, or perceive the Virgin Mary's image on a grilled cheese sandwich. Even in random data, we often find patterns, because—here's a curious fact of life—*random sequences often don't look random* (Falk et al., 2009; Nickerson, 2002, 2005). Flip a coin 50 times and you may be surprised at the streaks of heads or tails—much like supposed "hot" and "cold" streaks in basketball shooting and baseball hitting. In actual random sequences, patterns and streaks (such as repeating digits) occur more often than people expect (Oskarsson et al., 2009). That also makes it hard for people to generate random-like sequences. When embezzlers try to simulate random digits when specifying how much to steal, their nonrandom patterns can alert fraud experts (Poundstone, 2014).

Why are people prone to pattern-seeking? For most people, a random, unpredictable world is unsettling (Tullett et al., 2015). Making sense of our world relieves stress and helps us get on with daily living (Ma et al., 2017).

Some happenings, such as winning a lottery twice, seem so extraordinary that we find it difficult to conceive an ordinary, chance-related explanation. "But with a large enough sample," said statisticians Persi Diaconis and Frederick Mosteller (1989), "any outrageous thing is likely to happen." An event that happens to but 1 in 1 billion people every day occurs about 7 times a day, more than 2500 times a year.

"The really unusual day would be one where nothing unusual happens." — Statistician Persi Diaconis (2002)

The point to remember Our commonsense thinking is flawed due to three powerful tendencies—hindsight bias, overconfidence, and our tendency to perceive patterns in random events. But scientific inquiry can help us sift reality from illusion.

Play the role of a researcher using scientific inquiry to think smarter about random hot streaks in sports. Engage online with the activity ***How Would You Know If There Is a "Hot Hand" in Basketball?***

Psychological Science in a Post-Truth World

LOQ 1-2

Why are we so vulnerable to believing untruths?

In 2017, the Oxford English Dictionary's word of the year was *post-truth*—describing a modern culture where people's emotions and personal beliefs often override their acceptance of objective facts.

Consider two U.S. examples of such "truth decay"—of widely shared misinformation:

Belief *The crime rate is rising.* Every recent year, 7 in 10 adults told Gallup that there is more crime "than there was a year ago" (Swift, 2016).

Fact For several decades, both violent and property crime rates have been *falling.* In 2015, the violent crime rate was less than half the 1990 rate (BJS, 2017; Statista, 2017).

Belief *Many immigrants are criminals* (McCarthy, 2017). Memorable incidents feed this narrative. Stories of an immigrant murdering, burglarizing, or lying spread through social networks and news outlets. Such fears are commonplace not only in North America but also in Europe and Australia (Nunziata, 2015).

Fact Most immigrants are *not* criminals. Compared with native-born Americans, immigrants are 44 percent *less* likely to be imprisoned (CATO, 2017; Flagg, 2018, 2019). The same has been true in Italy, the United Kingdom, and elsewhere (Di Carlo et al., 2018).

Political party bias has distorted Americans' thinking. Psychologist Peter Ditto and his colleagues (2019a,b) reported that researchers have found "partisan bias in both liberals and conservatives, and at virtually identical levels." In the United States, a majority of Republicans believed unemployment had increased under Democratic President Barack Obama (it decreased), while a majority of strong Democrats believed inflation had worsened under Republican President Ronald Reagan (it improved) (Gelman, 2009; PPP, 2016). One study found that both U.S. Democrats and Republicans discriminate against the other-party candidates for college scholarships (Iyengar & Westwood, 2015). So, let none of us smugly think "Yes, but bias doesn't apply to *me*." Bias goes both ways.

U.S. Democrats and Republicans share concern about failures to separate fact from fiction. In his farewell address, President Barack Obama (2017) warned that without a "common baseline of facts," democracy is threatened: "We become so secure in our bubbles that we start accepting only information, whether it's true or not, that fits our opinions, instead of basing our opinions on the evidence that is out there." The late Republican Senator John McCain (2017) similarly expressed alarm about "the growing inability, and even unwillingness, to separate truth from lies."

Marty Bucella/Cartoon Stock

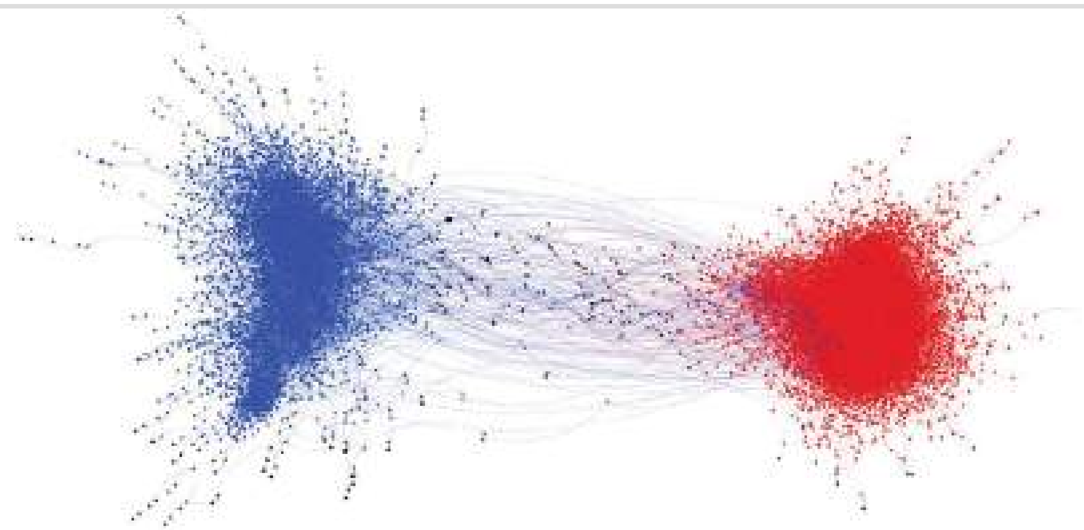
So why do post-truth era people so often, in the words of psychologist Tom Gilovich (1991), "know what isn't so?"

False news Some misinformation gets fed to us intentionally. It's "lies in the guise of news" (Kristof, 2017). In the 2016 U.S. election cycle, 6 percent of all Twitter-enabled news consumption was fake news (Grinberg et al., 2019). And made-up news persists. In one analysis of 126,000 stories tweeted by 3 million people, falsehoods—especially false political news—“diffused significantly farther, faster, deeper, and more broadly than the truth” (Vosoughi et al., 2018). The good news is that most people can often tell the difference between high- and low-quality information sources (Pennycook & Rand, 2019).

Repetition In experiments, statements become more believable when they are repeated (De keersmaecker et al., 2019). What we hear over and over—perhaps a made-up smear of a political opponent—gets remembered and comes to seem true (Fazio et al., 2015).

Availability of powerful examples In the media, “if it bleeds it leads.” Gruesome violence—a horrific murder, a mass killing, a plane crash—gets reported, with vivid images that color our judgments. No wonder Americans grossly overestimate their risk of being victimized by crime, terror, and plane crashes.

Group identity and the echo chamber of the like-minded Our social identities matter. Feeling good about our groups helps us feel good about ourselves. On social media we tend to friend people who think as we do (see **FIGURE 1.1**). We often read news sources that affirm our views and demonize news sources that do not.



Moral contagion in social networks William J. Brady, Julian A. Wills, John T. Jost, Joshua A. Tucker, Jay J.Van Bavel Proceedings of the National Academy of Sciences Jul 2017, 114 (28) 7313-7318; DOI:10.1073/pnas.1618923114

FIGURE 1.1 The meeting of like minds On social media, most people discuss contentious issues, such as gun control, same-sex marriage, and climate change, only with likeminded others. In this graph of politically charged Twitter activity, each node represents a user who sent a message; each line represents a user who retweeted another user. As we can see, users overwhelmingly sent messages to, and retweeted messages from, those who shared their liberal (blue) or conservative (red) ideology (Brady et al., 2017).

The good news is that we can build a real-truth world by embracing a scientific mindset. With a mix of curiosity, skepticism, and humility, we can adopt the spirit of critical thinking: To accept everything is to be gullible; to deny everything is to be a cynic.

> **"We have … become sloppier than ever: Tweet first, research later. Post first, rescind later. Guess first, confirm later."** — Luvvie Ajayi, *I'm Judging You: The Do-Better Manual,* 2016

"I'm sorry, Jeannie, your answer was correct, but Kevin shouted his incorrect answer over yours, so he gets the points."

Joe Dator The New Yorker Collection/The Cartoon Bank

To experience my [DM's] recap of some important, scientific thinking strategies, view the 3.5-minute animated video: ***Thinking Critically in Our "Post-Truth" World*** (also at tinyurl.com/PostTruthMyers).

The Scientific Method

The foundation of all science is a scientific attitude that combines *curiosity, skepticism,* and *humility*. Psychologists arm their scientific attitude with the *scientific method*—a self-correcting process for evaluating ideas with observation and analysis. Psychological science welcomes hunches and plausible-sounding theories. And it puts them to the test. If a theory works—if the data support its predictions—so much the better for that theory. If the predictions fail, the theory gets revised or rejected. When researchers submit their work to a scientific journal, *peer reviewers*—other scientists who are experts in that field—provide anonymous evaluations of a study's theory, originality, and accuracy. With this feedback in hand, the journal editor decides whether the research deserves publication.

Constructing Theories

LOQ 1-3

How do theories advance psychological science?

In everyday conversation, we often use *theory* to mean "mere hunch." Someone might, for example, discount evolution as "only a theory"—as if it were mere speculation. In science, a **theory** explains behaviors or events by offering ideas that organize observations. By using deeper principles to organize isolated facts, a theory summarizes and simplifies. As we connect the observed dots, a coherent picture emerges.

theory
an explanation using an integrated set of principles that organizes observations and predicts behaviors or events.

A theory of how sleep affects memory, for example, helps us organize countless sleep-related observations into a short list of principles. Imagine that we observe over and over that people with good sleep habits tend to answer questions correctly in class and do well at test time. We might therefore theorize that sleep improves memory. So far so good: Our principle neatly summarizes a list of observations about the effects of a good night's sleep.

Yet no matter how reasonable a theory may sound—and it does seem reasonable to suggest that sleep boosts memory—we must put it to the test. A good theory produces testable *predictions,* called **hypotheses**. Such predictions specify what results would support the theory and what results would disconfirm it. To test our theory about sleep effects on memory, we might hypothesize that when sleep deprived, people will remember less from the day before. To test that hypothesis, we might assess how well people remember course materials they studied either before a good night's sleep or before a shortened night's sleep (FIGURE 1.2). The results will either support our theory or lead us to revise or reject it.

hypothesis
a testable prediction, often implied by a theory.

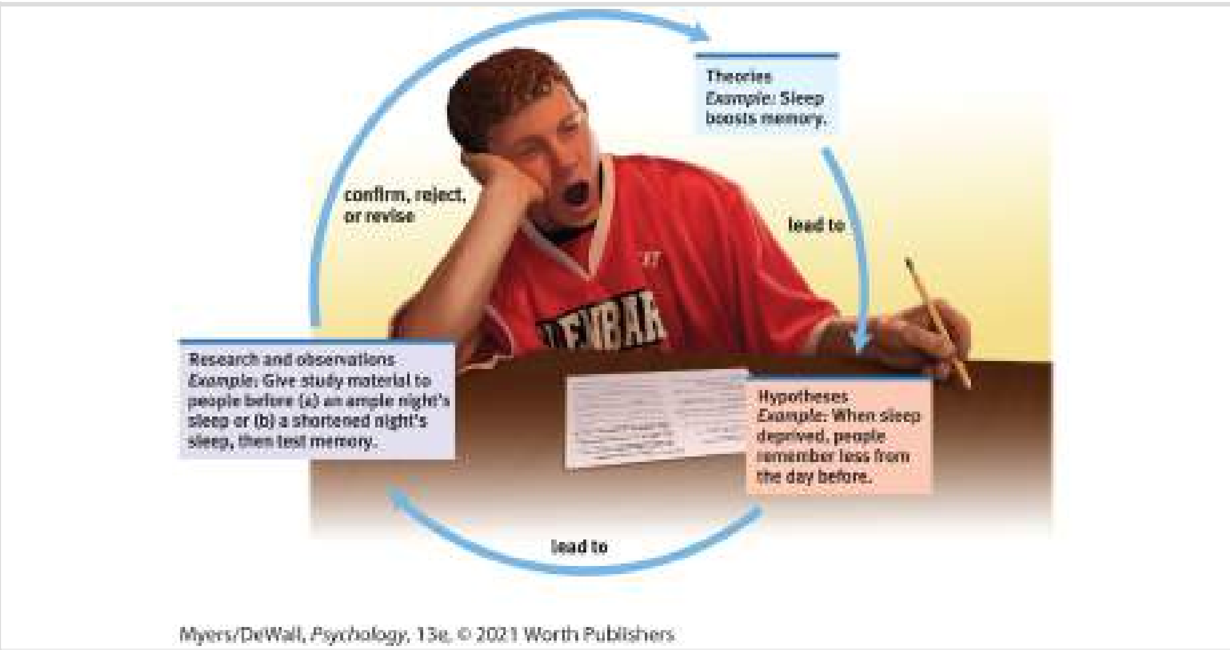


FIGURE 1.2 The scientific method A self-correcting process for asking questions and observing nature's answers.

Our theories can bias our observations. Having theorized that better memory springs from more sleep, we may see what we expect: We may perceive sleepy people's comments as less accurate. The urge to see what we expect is strong, both inside and outside the laboratory, as when people's views of climate change influence their interpretation of local weather events.

As a check on their own biases, psychologists report their research with precise, measurable **operational definitions** of research procedures and concepts. *Sleep deprived,* for example, may be defined as "at least 2 hours less" than the person's natural sleep. (Likewise, a study of "aggression" may observe how many pins you stab into a doll that represents a lab partner, or a study of "helping" may record dollars donated.) Using these carefully worded statements, others can **replicate** (repeat) the original observations with different participants, materials, and circumstances. If they get similar results, confidence in the finding's reliability grows. The first study of hindsight bias, for example, aroused psychologists' curiosity. Now, after many successful replications with differing people and questions, we feel sure of the phenomenon's power. Replication is confirmation.

operational definition
a carefully worded statement of the exact procedures (operations) used in a research study. For example, *human intelligence* may be operationally defined as what an intelligence test measures. (Also known as *operationalization.*)
replication
repeating the essence of a research study, usually with different participants in different situations, to see whether the basic finding can be reproduced.

Replication is an essential part of good science. Psychology experienced a "replication crisis" when recent multi-lab efforts to replicate as many as 100 studies produced mixed results—ranging from 36 percent to 85 percent of studies replicating (Camerer et al., 2018a; Klein et al., 2014, 2018; Open Science Collaboration, 2015). (None of these nonreproducible findings appear in this text.) Replication failures often result when samples are small, so psychologists increasingly study large samples of people (Camerer et al., 2018b; Sassenberg & Ditrich, 2019; Stanley et al., 2018). Bigger sample = a bigger chance of replication.

Today's psychological research is benefitting from more replications, more rigorous research methods, and more sharing of research data (Dougherty et al., 2018; Smaldino & McElreath, 2016; Zwaan et al., 2018). More and more psychologists use **preregistration** to publicly communicate their planned study design, hypotheses, data collection, and analyses (Nosek et al., 2018). (This openness and transparency also prevents later modifications, such as changing the hypotheses to fit the data.) There is still a place for *exploratory research* Investigators gather data and seek patterns that inspire theories, which can then be tested with *confirmatory research* (with preregistered hypotheses and preplanned analyses).

preregistration
publicly communicating planned study design, hypotheses, data collection, and analyses.

Explorations, replications, preregistrations, and the open sharing of raw data are enabling “Psychology’s Renaissance” of improved scientific practices (Motyl et al., 2017; Nelson et al., 2018).

> **“Failure to replicate is not a bug; it is a feature. It is what leads us along the path—the wonderfully twisty path—of scientific discovery.”** — Lisa Feldman Barrett, “Psychology Is Not in Crisis,” 2015

Psychological and medical science also harness the power of **meta-analysis.** Meta-analysis is a procedure for statistically synthesizing a body of scientific evidence. By combining the results of many studies, researchers avoid the problem of small samples and arrive at a bottom-line result.

meta-analysis
a statistical procedure for analyzing the results of multiple studies to reach an overall conclusion.

In the end, our theory will be useful if it (1) *organizes* observations and (2) implies *predictions* that anyone can use to check the theory or to derive practical applications. (Does people’s sleep predict their retention?) Eventually, our research may (3) stimulate further research that leads to a revised theory that better organizes and predicts.

As we will see next, we can test our hypotheses and refine our theories using *descriptive* methods (which describe behaviors, often through case studies, surveys, or naturalistic observations), *correlational* methods (which associate different factors), and *experimental* methods (which manipulate factors to discover their effects). To think critically about popular psychology claims, we need to understand these methods and know what conclusions they allow.

RETRIEVAL PRACTICE

RP-2 What does a good theory do?
RP-3 Why is replication important?

ANSWERS IN APPENDIX E

Description

LOQ 1-4

How do psychologists use case studies, naturalistic observations, and surveys to observe and describe behavior, and why is random sampling important?

The starting point of any science is description. In everyday life, we all observe and describe people, often drawing conclusions about why they think, feel, and act as they do. Psychologists do much the same, though more objectively and systematically, through

- ***case studies*** (in-depth analyses of individuals or groups).
- ***naturalistic observations*** (recording the natural behavior of many individuals).
- ***surveys* and interviews** (asking people questions).

THE CASE STUDY

Among the oldest research methods, the **case study** examines one individual or group in depth in the hope of revealing things true of us all. Some examples:

- ***Brain damage.*** Much of our early knowledge about the brain came from case studies of individuals who suffered particular impairments after damage to a certain brain region.
- ***Children's minds.*** Jean Piaget taught us about children's thinking after carefully observing and questioning only a few children.
- ***Animal intelligence.*** Studies of various animals, including a few chimpanzees, have revealed their capacity for understanding and language.

case study
a descriptive technique in which one individual or group is studied in depth in the hope of revealing universal principles.

Intensive case studies are sometimes very revealing, and they often suggest directions for further study.

But atypical individual cases may mislead us. Both in our everyday lives and in science, unrepresentative information can lead to mistaken judgments and false conclusions. Indeed, anytime a researcher mentions a finding *(Smokers die younger percent of men over are nonsmokers)* someone is sure to offer a contradictory anecdote *(Well, I have an uncle who smoked two packs a day and lived to be !)*.

Dramatic stories and personal experiences (even psychological case examples) command our attention and are easily remembered. Journalists understand that and often begin their articles with compelling stories. Stories move us. But stories can mislead. Which of the following do you find more memorable? (1) "In one study of 1300 dream reports concerning a kidnapped child, only 5 percent correctly envisioned the child as dead" (Murray & Wheeler, 1937). (2) "I know a man who dreamed his sister was in a car accident, and two days later she died in a head-on collision!" Numbers can be numbing, but *the plural of anecdote is not evidence*. A single story of someone who supposedly changed from gay to straight is not evidence that sexual orientation is a choice. As psychologist Gordon Allport (1954, p. 9) said, "Given a thimbleful of [dramatic] facts we rush to make generalizations as large as a tub."

Freud and Little Hans Sigmund Freud's case study of 5-year-old Hans' extreme fear of horses led Freud to his theory of childhood sexuality. He conjectured that Hans felt unconscious desire for his mother, feared castration by his rival father, and then transferred this fear into his phobia about being bitten by a horse. As Chapter 14 will explain, today's psychological science discounts Freud's theory of childhood sexuality but does agree that much of the human mind operates outside our conscious awareness.

The point to remember Individual cases can suggest fruitful ideas. What's true of all of us can be glimpsed in any one of us. But to find those general truths, we must employ other research methods.

RETRIEVAL PRACTICE

RP-4 We cannot assume that case studies always reveal general principles that apply to all of us. Why not?

ANSWERS IN APPENDIX E

See the ***Video: Case Studies*** for an animated tutorial.

NATURALISTIC OBSERVATION

A second descriptive method records responses in natural environments. These

naturalistic observations range from watching chimpanzee societies in the jungle, to videotaping and analyzing parent-child interactions in different cultures, to recording racial differences in students' self-seating patterns in a school lunchroom. In the digital age, naturalistic observations have increased—thanks to "big data" harvested from phone apps, social media, Google searches, and more.

naturalistic observation
a descriptive technique of observing and recording behavior in naturally occurring situations without trying to manipulate and control the situation.

Until recently, naturalistic observation was mostly "small science"—possible to do with pen and paper rather than fancy equipment and a big budget (Provine, 2012). But today's digital technologies have transformed naturalistic observations into big science. Want to keep track of how often people go to the gym, a café, or the library? All you need is access to their phone's global positioning system (GPS) (Harari et al., 2016). And new technologies—wearable cameras and fitness sensors, and internet-connected smart-home sensors—offer increasing possibilities for people to allow accurate recording of their activity, relationships, sleep, and stress (Nelson & Allen, 2018; Yokum et al., 2019).

The billions of people entering personal information online have also enabled big-data observations (without disclosing anyone's identity). One research team studied the ups and downs of human moods by counting positive and negative words in 504 million Twitter messages from 84 countries (Golder & Macy, 2011). As **FIGURE 1.3** shows, people seemed happier on weekends, shortly after waking, and in the evenings. (Are late Saturday evenings often a happy time for you, too?) Another study found that negative emotion (especially anger-related) words in 148 million tweets from 1347 U.S. counties predicted the counties' heart disease rates *better* than smoking and obesity rates (Eichstaedt et al., 2015). Google enables us to learn about the world, and people's Google use enables us to learn about them. For example, the words people search and the questions they ask can pinpoint a geographical area's level of racism and depression. But Google searches also reveal our universal human likeness—as illustrated by the word "pregnant" being searched in conjunction with the same food cravings across varied countries (Stephens-Davidowitz, 2017). Across the globe, we are kin beneath the skin.

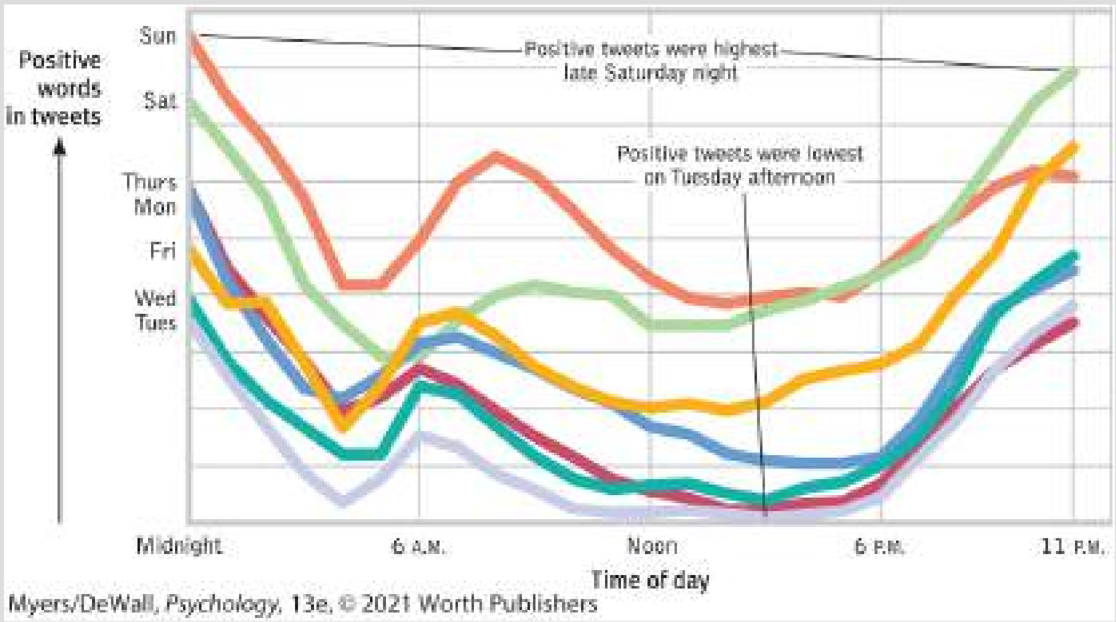


FIGURE 1.3 Twitter message moods, by time and by day This graph illustrates how, without knowing anyone's identity, researchers can use big data to study human behavior on a massive scale. It now is also possible to associate people's moods with, for example, their locations or with the weather, and to study the spread of ideas through social networks. (Data from Golder & Macy, 2011.)

Like the case study, naturalistic observation does not *explain* behavior. It *describes* it. Nevertheless, descriptions can be revealing. We once thought, for example, that only humans use tools. Then naturalistic observation revealed that chimpanzees sometimes insert a stick in a termite mound and withdraw it, eating the stick's load of termites. Such unobtrusive naturalistic observations paved the way for later studies of animal thinking, language, and emotion, which further expanded our understanding of our fellow animals. Thanks to researchers' observations, we know that chimpanzees and baboons use deception: Psychologists repeatedly saw one young baboon pretending to have been attacked by another as a tactic to get its mother to drive the other baboon away from its food (Whiten & Byrne, 1988).


MICHAEL NICHOLS/National Geographic Creative

A natural observer "Observations, made in the natural habitat," noted chimpanzee observer Jane Goodall (1998), "helped to show that the societies and behavior of animals are far more complex than previously supposed."

Naturalistic observations also illuminate human behavior. Here are two findings you might enjoy:

- ***A funny finding.*** We humans laugh 30 times more often in social situations than in solitary situations (Provine, 2001). (Have you noticed how seldom you laugh when alone?)
- ***Culture and the pace of life.*** Naturalistic observation also enabled Robert Levine and Ara Norenzayan (1999) to compare the pace of life—walking speed, accuracy of public clocks, and so forth—in 31 countries. Their conclusion: Life is fastest paced in Japan and Western Europe, and slower paced in economically less-developed countries.

Naturalistic observation offers interesting snapshots of everyday life, but it does so without controlling for all the factors that may influence behavior. It's one thing to observe the pace of life in various places, but another to understand what makes some people walk faster than others. Nevertheless, descriptions can be revealing: The starting point of any science is description.

RETRIEVAL PRACTICE

RP-5 What are the advantages and disadvantages of naturalistic observation?

ANSWERS IN APPENDIX E

See the ***Video: Naturalistic Observation*** for a helpful tutorial animation.

THE SURVEY

A **survey** looks at many cases, asking people to report their behavior or opinions. Questions about everything from sexual practices to political opinions are put to the public. Here are some recent survey findings:

- Compared at the same age with those born in the 1960s and 1970s, twice as many Millennials born in the 1990s reported having no sexual partners since age 18 (Twenge et al., 2017). Today's less attached young adults are experiencing what one writer termed a "sex recession" (Julian, 2018).
- 1 in 2 people across 24 countries reported believing in the "existence of intelligent alien civilizations in the universe" (Lampert, 2017).
- 68 percent of all humans—some 5.2 billion people—say that religion is important in their daily lives (from Gallup World Poll data analyzed by Diener et al., 2011).

survey
a descriptive technique for obtaining the self-reported attitudes or behaviors of a particular group, usually by questioning a representative, *random sample* of the group.

But asking questions is tricky. People may shade their answers in a socially desirable direction, such as by underreporting their cigarette consumption or overreporting their voting. And the answers often depend on how questions are worded and how respondents are chosen.

Wording Effects

Even small changes in the order or wording of questions can make a big difference

(**TABLE 1.1**). When U.S. White evangelical Christians were asked whether (1) “Humans have evolved over time” or (2) “Humans have existed in their present form since the beginning of time,” only 32 percent expressed a belief in evolution (Funk, 2019). But when asked whether (1) “Humans have evolved over time due to processes such as natural selection; God or a higher power had no part in this process”; (2) “Humans have evolved over time due to processes that were guided or allowed by God or a higher power”; or (3) “Humans have existed in their present form since the beginning of time,” more than twice as many—68 percent—expressed a belief in evolution. Because wording is such a delicate matter, critical thinkers will reflect on how the phrasing of a question might affect people’s expressed opinions.

TABLE 1.1 Survey Wording Effects

Garners More Approval	Garners Less Approval
“aid to the needy”	“welfare”
“affirmative action”	“preferential treatment”
“undocumented workers”	“illegal aliens”
“gun safety laws”	“gun control laws”
“revenue enhancers”	“taxes”
“enhanced interrogation”	“torture”

Random Sampling

In everyday thinking, we tend to generalize from samples we observe, especially vivid cases. An administrator who reads (a) a statistical summary of a professor’s student evaluations and (b) the vivid comments of two irate students may be influenced as much by the biased sample of two unhappy students as by the many favorable evaluations in the statistical summary. The temptation to succumb to the *sampling bias*—to generalize from a few vivid but unrepresentative cases—is nearly irresistible.

So how do you obtain a *representative sample?* Say you want to learn how students at your college or university feel about a proposed tuition increase. It’s often not possible to survey the whole group. How then could you choose a group that would represent the total student body? Typically, you would seek a **random sample**, in which every person in the entire **population** has an equal chance of being included in the sample group. You might number the names in the general student listing and use a random-number generator to pick your survey participants. (Sending each student a questionnaire wouldn't work because the conscientious people who returned it would not be a random sample.) Large representative samples are better than small ones, but a smaller representative sample of 100 is better than a larger unrepresentative sample of 500. You cannot compensate for an unrepresentative sample by simply adding more people.

random sample
a sample that fairly represents a population because each member has an equal chance of inclusion.
population
all those in a group being studied, from which random samples may be drawn. (*Note:* Except for national studies, this does *not* refer to a country's whole population.)

Political pollsters sample voters in national election surveys just this way. Without random sampling, large samples—such as from website polls—often give misleading results. But by using some 1500 randomly sampled people, drawn from all areas of a country, they can provide a remarkably accurate snapshot of the nation's opinions. Moreover, pollsters can assess the accuracy of their sampling by asking unrelated questions—such as whether the respondents live alone or are married—for which government statistics are available as benchmarks (Bialik, 2018). If the survey sample closely matches the national breakdown of people, so much the better.

Given polling's margin of error and last-minute voter swings, political polls are good but imperfect estimates of likely outcomes. Immediately before the 2016 U.S. presidential election, popular polling analysis website FiveThirtyEight.com gave candidate Hillary Clinton an estimated 71 percent chance of winning. When Donald Trump was then elected, many regarded the prediction as a failure. But consider: When a prediction model estimates a 71 percent chance for one candidate, that candidate should *lose* nearly one third of the time. (Imagine a weather forecast that predicts a 70 percent chance of rain. If it then *always* rained, that would be a flawed forecast.) One analysis of 30,000 general election political predictions in 45 countries between 1942 and 2017 summed it up: Contrary to popular belief, polls are pretty accurate (Jennings & Wlezien, 2018).

With very large samples, estimates become quite reliable. The letter *E* is estimated to represent 12.7 percent of the letters in written English. *E*, in fact, is 12.3 percent of the 925,141 letters in Melville's *Moby-Dick*, 12.4 percent of the 586,747 letters in Dickens' *A Tale of Two Cities*, and 12.1 percent of the 3,901,021 letters in 12 of Mark Twain's works (*Chance News*, 1997).

The point to remember Before accepting survey findings, think critically. Consider the sample. The best basis for generalizing is from a representative, random sample.

RETRIEVAL PRACTICE

RP-6 What is an unrepresentative sample, and how do researchers avoid it?

ANSWERS IN APPENDIX E

Correlation

LOQ 1-5

What does it mean when we say two things are correlated, and what are positive and negative correlations?

Describing behavior is a first step toward predicting it. Naturalistic observations and surveys often show us that one trait or behavior tends to coincide with another. In such cases, we say the two **correlate**. A statistical measure (the **correlation coefficient**) helps us figure out how closely two things vary together, and thus how well either one *predicts* the other. Knowing how much aptitude test scores *correlate* with school success tells us how well the scores *predict* school success.

correlation
a measure of the extent to which two factors vary together, and thus of how well either factor predicts the other.
correlation coefficient
a statistical index of the relationship between two things (from −1.00 to +1.00).

Throughout this book, we often ask how strongly two **variables** are related: For example, how closely related are the personality test scores of identical twins? How well do intelligence test scores predict career achievement? How much do people's depressive symptoms predict their anxiety? In such cases, **scatterplots** can be very revealing.

variable
anything that can vary and is feasible and ethical to measure.
scatterplot
a graphed cluster of dots, each of which represents the values of two variables. The slope of the points suggests the direction of the relationship between the two variables. The amount of scatter suggests the strength of the correlation (little scatter indicates high correlation).

Each dot in a scatterplot represents the values of two variables. The three scatterplots in **FIGURE 1.4** illustrate the range of possible correlations from a perfect positive to a perfect negative. (Perfect correlations rarely occur in the real world.) A correlation is positive if two sets of scores, such as for height and weight, tend to rise or fall together.

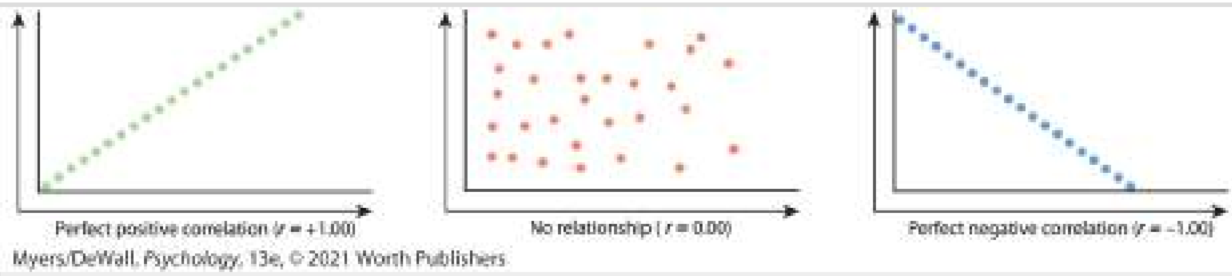


FIGURE 1.4 Scatterplots, showing patterns of correlation Correlations—abbreviated *r*—can range from +1.00 (scores for one variable increase in direct proportion to scores for another), to 0.00 (no relationship), to −1.00 (scores for one variable decrease precisely as scores for the other rise).

Saying that a correlation is “negative” says nothing about its strength. A correlation is negative if two sets of scores relate inversely, one set going up as the other goes down. The correlation between people’s height and the distance from their head to the ceiling is strongly (perfectly, in fact) negative.

Statistics can help us understand what we might miss with casual observation. To demonstrate, consider the responses of 2291 Czech and Slovakian volunteers who were asked by Jakub Polák and his colleagues (2019) to rate, on a 1 to 7 scale, their *fear* and *disgust* related to each of 24 animals. With all the relevant data right in front of you (**TABLE 1.2**), can you tell whether the correlation between participants’ fear and their disgust is positive, negative, or close to zero?

TABLE 1.2 People’s Fear and Disgust Responses to Various Animals

Animal	Average Fear	Average Disgust
Ant	2.12	2.26
Bat	2.11	2.01
Bull	3.84	1.62
Cat	1.24	1.17
Cockroach	3.10	4.16
Dog	2.25	1.20
Fish	1.15	1.38
Frog	1.84	2.48
Grass snake	3.32	2.47
Horse	1.82	1.11
Lizard	1.46	1.46
Louse	3.58	4.83
Maggot	2.90	4.49
Mouse	1.62	1.78
Panda	1.57	1.17
Pigeon	1.48	2.01
Rat	2.11	2.25
Rooster	1.78	1.34
Roundworm	3.49	4.79
Snail	1.15	1.69
Spider	4.39	4.47
Tapeworm	3.60	4.83
Viper	4.34	2.83
Wasp	3.42	2.84

When comparing the columns in Table 1.2, most people don’t detect a relationship between fear and disgust. In fact, the correlation in this imaginary example is positive ($r = +.72$), as we can see if we display the data as a scatterplot (**FIGURE 1.5**).

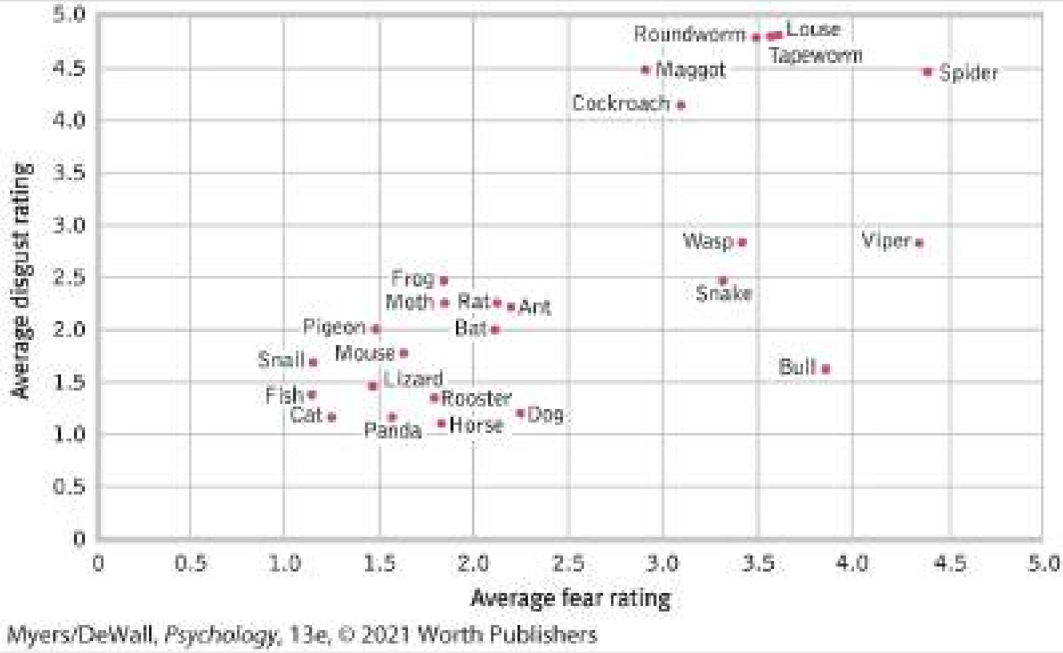


FIGURE 1.5 Scatterplot for fear and disgust felt toward 24 animals This display of average self-reported fear and disgust (each represented by a data point) reveals an upward slope, indicating a positive correlation. The considerable scatter of the data indicates the correlation is much lower than +1.00.

If we don't easily recognize a strong relationship when data are presented systematically, as in Table 1.2, how much less likely are we to notice them in everyday life? To see what is right in front of us, we sometimes need statistical illumination. We can easily see evidence of gender discrimination when given statistically summarized information about job level, seniority, performance, gender, and salary. But we often see no discrimination when the same information dribbles in, case by case (Twiss et al., 1989). Thinking like a psychological scientist helps us value everyone equally—not just those who catch our attention.

The point to remember A correlation coefficient helps us see the world more clearly by revealing the extent to which two things relate.

RETRIEVAL PRACTICE

RP-7 Indicate whether each association is a positive correlation or a negative correlation.

1. The more husbands viewed internet pornography, the worse their marital relationships (Muusses et al., 2015).

   ____

2. The more time teen girls spend absorbed with online social media, the more at risk they are for depression and

suicidal thoughts (Kelly et al., 2018; Twenge & Campbell, 2019). ______

3. The longer children were breast-fed, the greater their later academic achievement (Horwood & Fergusson, 1998). ______
4. The more leafy vegetables older adults eat, the less their mental decline over the ensuing 5 years (Morris et al., 2018). ______

ANSWERS IN APPENDIX E

For an animated tutorial on correlations, engage online with ***Concept Practice: Positive and Negative Correlations.*** See also the ***Video: Correlational Studies*** for another helpful tutorial animation.

ILLUSORY CORRELATIONS AND REGRESSION TOWARD THE MEAN

LOQ 1-6

What are *illusory correlations,* and what is *regression toward the mean?*

Correlations not only make clear the relationships we might otherwise miss; they also keep us from falsely observing nonexistent relationships. When we believe there is a relationship between two things, we are likely to notice and recall instances that confirm our belief. If we believe that dreams forecast actual events, we may notice and recall confirming instances more than disconfirming instances. The result is an **illusory correlation**.

illusory correlation
perceiving a relationship where none exists, or perceiving a stronger-than-actual relationship.

Illusory correlations can feed an illusion of control—that chance events are subject to our personal control. Gamblers, remembering their lucky rolls, may come to believe they can influence the roll of the dice by again throwing gently for low numbers and hard for high numbers. The illusion that uncontrollable events correlate with our actions is also fed by a statistical phenomenon called **regression toward the mean**. Average results are more typical than extreme results. Thus, after an unusual event, things tend to return toward their average level; extraordinary happenings tend to be followed by more ordinary ones. Students who score much lower or higher on an exam than they usually do are likely, when retested, to return to their average. Extrasensory perception (ESP) test subjects who defy chance when first tested nearly always lose their "psychic powers" when retested.

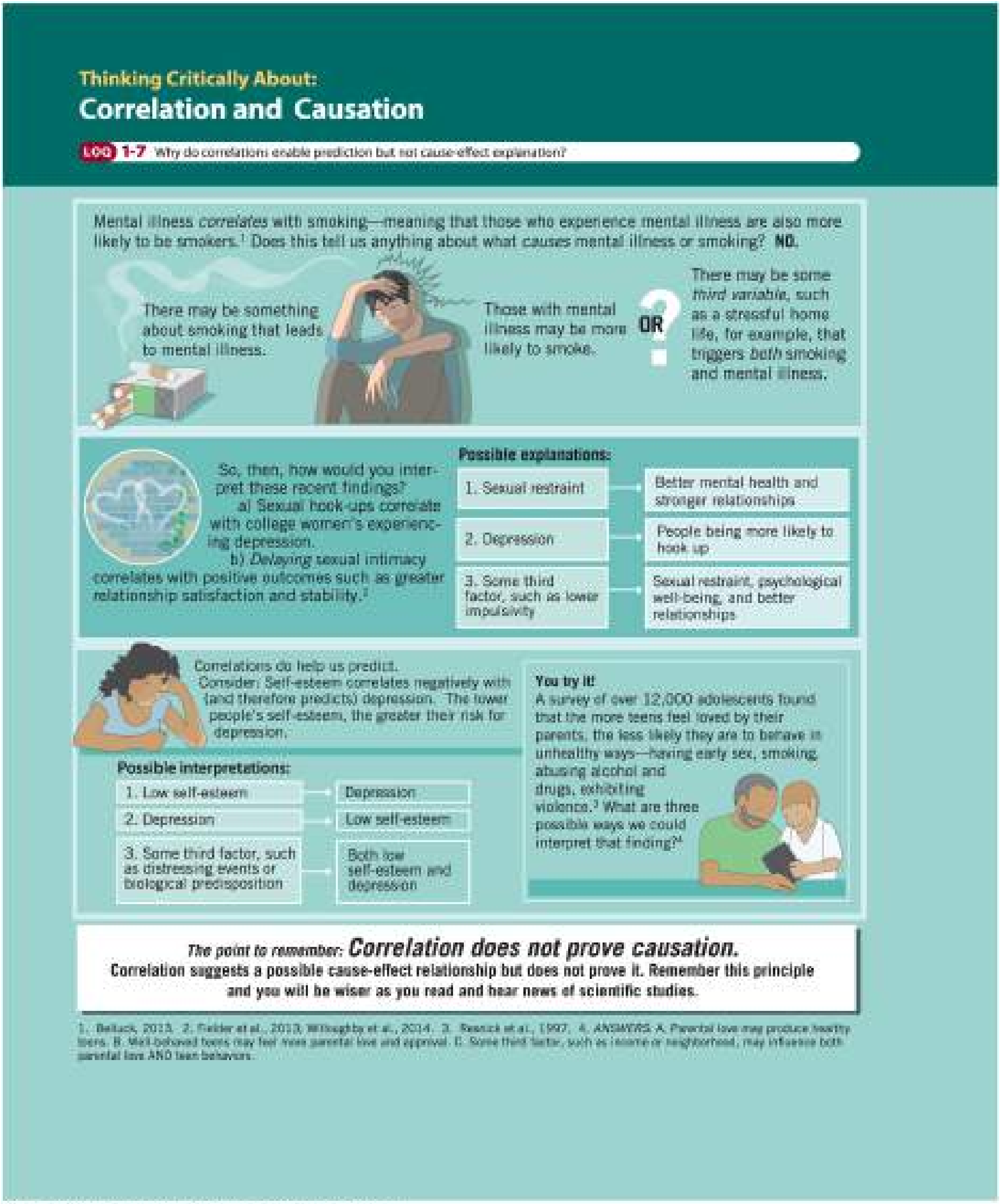
regression toward the mean
the tendency for extreme or unusual scores or events to fall back (regress) toward the average.

Failure to recognize regression is the source of many superstitions and of some ineffective practices—in sports coaching and in the workplace, for example. After berating an employee (or athlete) for poorer-than-usual performance, a manager may—when the employee regresses to normal—feel rewarded by the "improvement." After lavishing praise for an exceptionally fine performance, the manager may be disappointed when the employee's behavior again migrates back toward his or her average. Ironically, then, regression toward the average can mislead us into feeling rewarded after criticizing others ("That criticism really made him work harder!") and feeling punished after praising them ("All those compliments made her slack off!") (Tversky & Kahneman, 1974).

> **"Once you become sensitized to it, you see regression everywhere."** — Psychologist Daniel Kahneman (1985)

The point to remember When a fluctuating behavior returns to normal, fancy explanations for why it does so are often wrong. Regression toward the mean is probably at work.

Moreover, although correlational research helpfully reveals relationships, it does not explain them. If teen social media use correlates with (predicts) teen risk of depression, that may—or may not—indicate that social media use affects depression risk. Other explanations are possible (see Thinking Critically About: Correlation and Causation).



Myers/DeWall, *Psychology*, 13e, © 2021 Worth Publishers

RETRIEVAL PRACTICE

RP-8 You hear the school basketball coach telling her friend that she rescued her team's winning streak by yelling at the players after an unusually bad first half. What is another explanation of why the team's performance improved?

RP-9 Length of marriage positively correlates with hair loss in men. Does this mean that marriage causes men to lose their hair (or that balding men make better husbands)?

Nancy Brown/Getty Images

ANSWERS IN APPENDIX E

Experimentation

LOQ 1-8

What are the characteristics of experimentation that make it possible to isolate cause and effect?

Happy are they, remarked the Roman poet Virgil, "who have been able to perceive the causes of things." How might psychologists sleuth out the causes in correlational studies, such as the small correlation between teen girls' social media use and their risk of depression and self-harm?

EXPERIMENTAL MANIPULATION

Our sleuthing starts with two plain facts:

1. Beginning in 2010, worldwide smart phone and social media use mushroomed.
2. Simultaneously, Canadian, American, and British teen girls' rates of depression, anxiety, self-harm, and suicide also mushroomed (Mercado et al., 2017; Morgan,

2017; Statistics Canada, 2016).

What do such findings mean? Is there a cause-effect connection? If so, should parents limit their middle schoolers' time on Instagram or Snapchat? Even big correlational data from a million teens couldn't tell us. The answers are being debated and the data are mixed. Moving beyond the simple correlation, one research summary noted that, in six of eight *longitudinal* (over time) studies, teens' current social media use predicted future mental health issues (Haidt, 2019). Even so, to identify cause and effect, researchers must **experiment**. Experiments enable researchers to isolate the effects of one or more factors by (1) *manipulating the factors of interest* and (2) *holding constant (controlling") other factors.* To do so, they often create an **experimental group** in which people receive the treatment (such as reduced screen time), and a contrasting **control group** in which they do not.

experiment
a research method in which an investigator manipulates one or more factors (independent variables) to observe the effect on some behavior or mental process (the dependent variable). By *random assignment* of participants, the experimenter aims to control other relevant factors.
experimental group
in an experiment, the group exposed to the treatment, that is, to one version of the independent variable.
control group
in an experiment, the group *not* exposed to the treatment; contrasts with the experimental group and serves as a comparison for evaluating the effect of the treatment.

To minimize any preexisting differences between the two groups, experimenters **randomly assign** people to each condition. Random assignment—whether with a random numbers table or flip of the coin—effectively equalizes the two groups. If one-third of the volunteers for an experiment can wiggle their ears, then about one-third of the people in each group will be ear wigglers. So, too, with age, attitudes, and other characteristics, which will be similar in the experimental and control groups. Thus, if the groups differ at the experiment's end, we can surmise that the treatment had an effect. (Note the difference between random *sampling*—which creates a representative survey sample—and random *assignment,* which equalizes the experimental and control groups.)

random assignment
assigning participants to experimental and control groups by chance, thus minimizing preexisting differences between the

different groups.

So, what do *experiments* reveal about the relationship between girls' social media use and their risk of depression and self-harm? While there have been few actual experiments that vary social media use, one identified nearly 1700 people who agreed to deactivate their Facebook account for four weeks (Allcott et al., 2019). Compared with people in the control group, those randomly assigned to the deactivation group spent more time watching TV and socializing with friends and family—and they reported lower depression, and greater happiness and satisfaction with their lives (and less post-experiment Facebook use). Less Facebook time meant a happier life.

The debate over the effects of prolonged social media use is ongoing. For now, most researchers agree that unlimited teen social media use poses a modest mental health risk. With more large-scale correlational and longitudinal studies, and further experimentation, researchers will refine this tentative conclusion.

The point to remember Correlational studies, which uncover naturally occurring relationships, are complemented by experiments, which manipulate a factor to determine its effect.

See the ***Video: Random Assignment*** for a tutorial animation.

PROCEDURES AND THE PLACEBO EFFECT

Consider, then, how we might assess therapeutic interventions. Our tendency to seek new remedies when we are ill or emotionally down can produce misleading testimonies. If three days into a cold we start taking zinc tablets and find our cold symptoms lessening, we may credit the pills rather than the cold naturally subsiding. In the 1700s, bloodletting *seemed* effective. People sometimes improved after the treatment; when they didn't, the practitioner inferred the disease was too advanced to be reversed. So, whether or not a remedy is truly effective, enthusiastic users will probably endorse it. To determine its effect, we must control for other factors.

And that is precisely how new drugs and new methods of psychological therapy are evaluated (Chapter 16). Investigators randomly assign participants in these studies to research groups. One group receives a pseudotreatment—an inert *placebo* (perhaps a pill with no drug in it). The other group receives a treatment, such as an antidepressant medication. (You can think of the placebo versus the actual drug as "trick or treatment.") The participants are often *blind* (uninformed) about what treatment, if any, they are receiving. If the study is using a **double-blind procedure,** neither the participants nor those who administer the drug and collect the data will know which group is receiving the treatment.

double-blind procedure
an experimental procedure in which both the research participants and the research staff are ignorant (blind) about whether the research participants have received the treatment or a placebo. Commonly used in drug-evaluation studies.

In double-blind studies, researchers check a treatment's actual effects apart from the participants' and the staff's belief in its healing powers. Just *thinking* you are getting a treatment can boost your spirits, relax your body, and relieve your symptoms. This **placebo effect** is well documented in reducing pain, depression, anxiety, and auditory hallucinations in schizophrenia (Dollfus et al., 2016; Kirsch, 2010). Athletes have run faster when given a supposed performance-enhancing drug (McClung & Collins, 2007). Decaf-coffee drinkers have reported increased vigor and alertness when they thought their brew had caffeine in it (Dawkins et al., 2011). People have felt better after receiving a phony mood-enhancing drug (Michael et al., 2012). And the more expensive the placebo, the more "real" it seems to us—a fake pill that costs $2.50 worked better than one costing 10 cents (Waber et al., 2008). To know how effective a therapy really is, researchers must control for a possible placebo effect.

placebo [pluh-SEE-bo; Latin for "I shall please"] effect
experimental results caused by expectations alone; any effect on behavior caused by the administration of an inert substance or condition, which the recipient assumes is an active agent.

"If I don't think it's going to work, will it still work?"

RETRIEVAL PRACTICE

RP-10 What measures do researchers use to prevent the *placebo effect* from confusing their results?

ANSWERS IN APPENDIX E

INDEPENDENT AND DEPENDENT VARIABLES

Here is an even more potent example: The drug Viagra was approved for use after 21 clinical trials. One trial was an experiment in which researchers randomly assigned 329 men with erectile disorder to either an experimental group (Viagra takers) or a control group (placebo takers given an identical-looking pill). The procedure was double-blind—neither the men taking the pills nor the person giving them knew what participants were receiving. The result: At peak doses, 69 percent of Viagra-assisted attempts at intercourse were successful, compared with 22 percent for men receiving the placebo (Goldstein et al., 1998). Viagra performed.

This simple experiment manipulated just one factor: the drug (Viagra versus no Viagra). We call this experimental factor the **independent variable** because we can vary it *independently* of other factors, such as the men's age, weight, and personality. Other factors that can potentially influence a study's results are called **confounding variables**. Random assignment controls for possible confounding variables.

independent variable
in an experiment, the factor that is manipulated; the variable whose effect is being studied.
confounding variable
in an experiment, a factor other than the factor being studied that might influence a study's results.

Experiments examine the effect of one or more independent variables on some measurable behavior, called the **dependent variable** because it can vary *depending* on what takes place during the experiment. Both variables are given precise *operational definitions,* which specify the procedures that manipulate the independent variable (the exact drug dosage and timing in this study) or measure the dependent variable (the men's responses to questions about their sexual performance). These definitions offer a level of precision that enables others to replicate the study. (See **FIGURE 1.6** for the Facebook experiment design.)

dependent variable
in an experiment, the outcome that is measured; the variable that may change when the independent variable is manipulated.

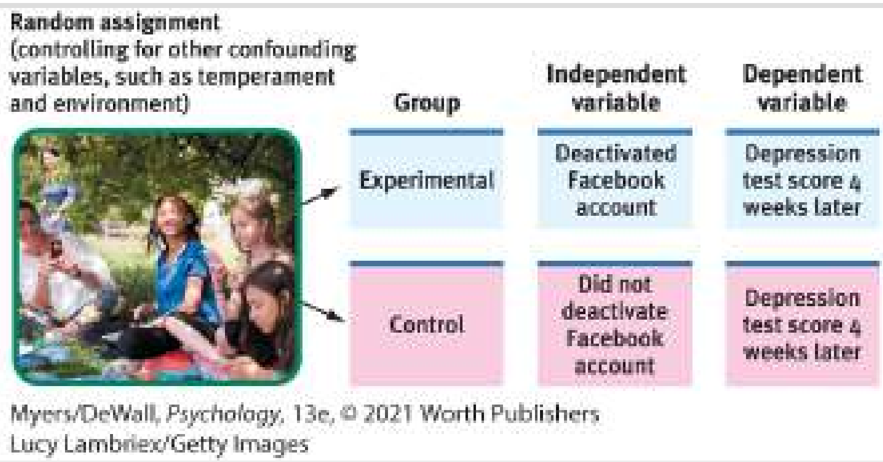


FIGURE 1.6 Experimentation To discern causation, psychologists control for confounding variables by randomly assigning some participants to an experimental group, others to a control group. Measuring the dependent variable (depression test score) will determine the effect of the independent variable (social media exposure).

A similar experiment on a drug approved to increase women's sexual arousal produced a result described as, um, anticlimactic—an additional "half of one satisfying sexual encounter a month" (Ness, 2016; Tavernise, 2016).

Let’s pause to check your understanding using a simple psychology experiment: To test the effect of perceived ethnicity on the availability of rental housing, Adrian Carpusor and William Loges (2006) sent identically worded email inquiries to 1115 Los Angeles–area landlords. The researchers varied the ethnic connotation of the sender’s name and tracked the percentage of positive replies (invitations to view the apartment in person). “Patrick McDougall,” “Said Al-Rahman,” and “Tyrell Jackson” received, respectively, 89 percent, 66 percent, and 56 percent invitations. In this experiment, what was the independent variable? The dependent variable?[3]

> **“[We must guard] against not just racial slurs, but … against the subtle impulse to call Johnny back for a job interview, but not Jamal.”** — U.S. President Barack Obama, eulogy for state senator and church-shooting victim Clementa Pinckney, June 26, 2015

Experiments can also help us evaluate social programs. Do early childhood education programs boost impoverished children’s chances for success? What are the effects of different antismoking campaigns? Do school sex-education programs reduce teen pregnancies? To answer such questions, we can experiment: If an intervention is welcomed but resources are scarce, we could use a lottery to randomly assign some people (or regions) to experience the new program and others to a control condition. If later the two groups differ, the intervention’s effect will be supported (Passell, 1993).

Let’s recap. A *variable* is anything that can vary (infant nutrition, intelligence, social media exposure—anything within the bounds of what is feasible and ethical to measure). Experiments aim to *manipulate* an *independent* variable, *measure* a *dependent* variable, and *control confounding* variables. An experiment has at least two different conditions: an *experimental condition* and a *comparison* or *control condition. Random assignment* works to minimize preexisting differences between the groups before any treatment effects occur. In this way, an experiment tests the effect of at least one independent variable (what we manipulate) on at least one dependent variable (the outcome we measure).

RETRIEVAL PRACTICE

RP-11 By using *random assignment,* researchers are able to control for ________________, which are other factors besides the independent variable(s) that may influence research results.

RP-12 Match the term on the left (i through iii) with the description on the right (a through c).

i. Double-blind procedure	a. helps researchers generalize from a small set of survey responses to a larger population
ii. Random sampling	b. helps minimize preexisting differences between experimental and control groups
iii. Random assignment	c. controls for the placebo effect; neither researchers nor participants know who receives the real treatment

RP-13 Why, when testing a new drug to control blood pressure, would we learn more about its effectiveness from giving it to half the participants in a group of 1000 than to all 1000 participants?

ANSWERS IN APPENDIX E

See the ***Videos: Experiments*** and ***Confounding Variables*** for helpful tutorial animations.

Research Design

LOQ 1-9

How would you know which research design to use?

Throughout this book, you will read about amazing psychological science discoveries. But how do psychological scientists choose research methods and design their studies in ways that provide meaningful results? Understanding how research is done—how testable questions are developed and studied—is key to appreciating all of psychology. **TABLE 1.3** compares the features of psychology's main research methods. In later chapters, you will read about other research designs, including *twin studies* (Chapter 4) and *cross-sectional* and *longitudinal research* (Chapter 10).

TABLE 1.3 Comparing Research Methods

Research Method	Basic Purpose	How Conducted	What Is Manipulated	Weaknesses
Descriptive	To observe and record behavior	Do case studies, naturalistic observations, or surveys	Nothing	No control of variables; single cases may be misleading
Correlational	To detect naturally occurring relationships; to assess how well one variable predicts another	Collect data on two or more variables; no manipulation	Nothing	Cannot specify cause and effect
Experimental	To explore cause and effect	Manipulate one or more factors; use random assignment	The independent variable(s)	Sometimes not feasible; results may not generalize to other contexts; not ethical to manipulate certain variables

In psychological research, no questions are off limits, except untestable (or unethical) ones: Does free will exist? Are people born evil? Is there an afterlife? Psychologists can't test those questions. But they *can* test whether free will beliefs, aggressive personalities, and a belief in life after death influence how people think, feel, and act (Dechesne et al., 2003; Shariff et al., 2014; Webster et al., 2014).

Having chosen their question, psychologists then select the most appropriate research design—*experimental, correlational, case study, naturalistic observation, twin study, longitudinal,* or *cross-sectional*—and determine how to set it up most effectively. They consider how much money and time are available, ethical issues, and other limitations. For example, it wouldn't be ethical for a researcher studying child development to use the experimental method and randomly assign children to loving versus punishing homes.

Next, psychological scientists decide how to measure the behavior or mental process being studied. For example, researchers studying aggressive behavior could measure participants' willingness to blast a stranger with supposed intense noise.

Researchers want to have confidence in their findings, so they carefully consider confounding variables—factors other than those being studied that may affect their interpretation of results.

Psychological research is a creative adventure. Researchers *design* each study, *measure* target behaviors, *interpret* results, and learn more about the fascinating world of behavior and mental processes along the way.

ASK YOURSELF

If you could conduct a study on any psychological question, which question would you choose? How would you design the study?

To help you build your understanding, your critical thinking, and your *scientific literacy skills,* we created online research activities. In these *How Would You Know?* activities, you get to play the role of researcher, making choices about the best ways to test interesting questions. Some examples: How Would You Know If Having Children Relates to Being Happier?, How Would You Know If a Cup of Coffee Can Warm Up Relationships?, and How Would You Know If People Can Learn to Reduce Anxiety?

To review and test your understanding of research methods, engage online with ***Concept Practice: Psychology's Research Methods*** and ***The Language of Experiments,*** and the interactive ***Topic Tutorial: PsychSim6, Understanding Psychological Research.*** For a 9.5-minute video synopsis of psychology's scientific research strategies, see the ***Video: Research Methods.***

Predicting Everyday Behavior

LOQ 1-10

How can simplified laboratory conditions illuminate everyday life?

When you see or hear about psychological research, do you ever wonder whether people's behavior in the lab will predict their behavior in everyday life? Does detecting the blink of a faint red light in a dark room say anything useful about flying an airplane at night? After viewing a violent, sexually explicit film, does a man's increased willingness to push buttons that he thinks will deliver a noise blast to a woman really say anything about whether viewing violent pornography makes a man more likely to abuse a woman?

Before you answer, consider: The experimenter *intends* the laboratory environment to be a simplified reality—one that simulates and controls important features of everyday life. Just as a wind tunnel lets airplane designers re-create airflow forces under controlled conditions, a laboratory experiment lets psychologists re-create psychological forces under controlled conditions.

An experiment's purpose is not to re-create the exact behaviors of everyday life, but to test *theoretical principles* (Mook, 1983). In aggression studies, deciding whether to push a button that delivers a noise blast may not be the same as slapping someone in the face, but the principle is the same. *It is the resulting principles—not the specific findings—that help explain everyday behaviors.*

When psychologists apply laboratory research on aggression to actual violence, they are applying theoretical principles of aggressive behavior, principles they have refined through many experiments. Similarly, it is the principles of the visual system, developed from experiments in artificial settings (such as looking at red lights in the dark), that researchers apply to more complex behaviors such as night flying. And many investigations show that principles derived in the laboratory do typically generalize to the everyday world (Mitchell, 2012).

The point to remember Psychological science focuses less on specific behaviors than on revealing general principles that help explain many behaviors.

Psychology's Research Ethics

LOQ 1-11

Why do psychologists study animals, and what ethical research guidelines safeguard human and animal welfare? How do psychologists' values influence what they study and how they apply their results?

We have reflected on how a scientific approach can restrain biases. We have seen how case studies, naturalistic observations, and surveys help us describe behavior. We have also noted that correlational studies assess the association between two factors, showing how well one predicts another. We have examined the logic that underlies experiments, which use control conditions and random assignment of participants to isolate the causal effects of an independent variable on a dependent variable.

Yet, even knowing this much, you may still be approaching psychology with a mixture of curiosity and apprehension. So before we plunge in, let’s entertain some common questions about psychology’s ethics and values.

See the ***Video: Research Ethics*** for a helpful tutorial animation.

Studying and Protecting Animals

Many psychologists study nonhuman animals because they find them fascinating. They want to understand how different species learn, think, and behave. Psychologists also study animals to learn about people. We humans are not *like* animals; we *are* animals, sharing a common biology. Animal experiments have therefore led to treatments for human diseases—insulin for diabetes, vaccines to prevent polio and rabies, transplants to replace defective organs.

Humans are complex. But some of the same processes by which we learn are present in other animals, even sea slugs and honeybees. The simplicity of the sea slug’s nervous system is precisely what makes it so revealing of the neural mechanisms of learning. Ditto for the honeybee, which resembles us humans in how it learns to cope with stress (Dinges et al., 2017).

Sharing such similarities, should we not respect our animal relatives? The animal protection movement protests the use of animals in psychological, biological, and medical research. “We cannot defend our scientific work with animals on the basis of the similarities between them and ourselves and then defend it morally on the basis of differences,” noted Roger Ulrich (1991). In U.S. national surveys, half of adults oppose and half favor “the use of animals in scientific research”—with support greater among those most informed about science (Strauss, 2018).

> **"Rats are very similar to humans except that they are not stupid enough to purchase lottery tickets."** — Dave Barry, July 2, 2002

Out of this heated debate, two issues emerge. The basic one is whether it is right to place the well-being of humans above that of other animals. In experiments on stress and cancer, is it right that mice get tumors in the hope that people might not? Should some monkeys be exposed to an HIV-like virus in the search for an AIDS vaccine? Humans raise and slaughter 56 billion animals a year (Thornton, 2019). Is our use and consumption of other animals as natural as the behavior of carnivorous hawks, cats, and whales?

> **"Please do not forget those of us who suffer from incurable diseases or disabilities who hope for a cure through research that requires the use of animals."** — Psychologist Dennis Feeney (1987)

For those who give human life top priority, a second question emerges: What safeguards should protect the well-being of animals in research? One survey of animal researchers gave an answer. Some 98 percent supported government regulations protecting primates, dogs, and cats, and 74 percent also supported regulations providing for the humane care of rats and mice (Plous & Herzog, 2000). Many professional associations and funding agencies already have such guidelines. British Psychological Society (BPS) guidelines call for housing animals under reasonably natural living conditions, with companions for social animals (Lea, 2000). American Psychological Association (APA) guidelines state that researchers must provide "humane care and healthful conditions" and that testing should "minimize discomfort" (APA, 2012). The European Parliament also mandates standards for animal care and housing (Vogel, 2010). Most universities screen research proposals, often through an animal care ethics committee, and laboratories are regulated and inspected.

> **"The greatness of a nation can be judged by the way its animals are treated."** — Mahatma Gandhi, 1869–1948

Animals have themselves benefited from animal research. One Ohio team of research psychologists measured stress hormone levels in samples of millions of dogs brought each year to animal shelters. They devised handling and stroking methods to reduce stress and ease the dogs' transition to adoptive homes (Tuber et al., 1999). Other studies have helped improve care and management in animals' natural habitats. By revealing our behavioral kinship with animals and the remarkable intelligence of chimpanzees, gorillas, and other animals, experiments have also led to increased empathy and protection for them. At its best, a psychology concerned for humans and sensitive to animals serves the welfare of both.

MARY ALTAFFER/AP Images

Animal research benefiting animals Psychologists have helped zoos enrich animal environments—for example, by giving animals more choices to reduce the *learned helplessness* of captivity (Kurtycz, 2015; Weir, 2013). Thanks partly to research on the benefits of novelty, control, and stimulation, these gorillas are enjoying an improved quality of life in New York's Bronx Zoo.

Studying and Protecting Humans

What about human participants? Does the image of white-coated scientists seeming to deliver electric shocks trouble you? Actually, most psychological studies are free of such stress. Blinking lights, flashing words, and pleasant social interactions are more common.

Occasionally, researchers do temporarily stress or deceive people, but only when they believe it is essential to a justifiable end, such as understanding and controlling violent behavior or studying mood swings. Some experiments won't work if participants know everything beforehand. (Wanting to be helpful, the participants might try to confirm the researcher's predictions.)

The ethics codes of the APA and Britain's BPS urge researchers to (1) obtain potential participants' **informed consent** to take part, (2) protect participants from greater-than-usual harm and discomfort, (3) keep information about individual participants confidential, and (4) fully **debrief** people (explain the research afterward, including any temporary deception). To enforce these ethical standards, universities and research organizations have *Institutional Review Boards* that screen research proposals and safeguard "the rights, welfare and well-being of human research participants" (NIEHS, 2019).

informed consent
giving potential participants enough information about a study to enable them to choose whether they wish to participate.
debriefing
the postexperimental explanation of a study, including its purpose and any deceptions, to its participants.

Ensuring Scientific Integrity

In science, as in everyday life, mistakes happen. When data get accidentally miscomputed or misreported, that's forgivable and correctable. What's not acceptable —and will get a scientist banished from the profession—is fraud. Indeed, leading scientists cite honesty as the most important scientific value, followed by curiosity and perseverance (*Nature,* 2016). To seek career advancement by plagiarizing another's words or ideas, or to make up data, is to risk finding one's career ended. Such was the rare case when a Dutch psychologist fabricated data that made it into 58 research articles—fakery that was sniffed out by alert colleagues (Retraction Watch, 2015).

Fake science also has the potential to cause great harm. This happened in 1998 when a now-disbarred British physician, Andrew Wakefield, published an article in the prestigious *Lancet,* reporting a dozen cases in which British children given the measles, mumps, and rubella (MMR) vaccine supposedly developed autism afterward. Other studies failed to reproduce the finding (replication matters!) (Hviid et al., 2019). An investigation revealed a fraud—with falsified data—and the journal retracted the report (Godlee, 2011). Alas, by then the widely publicized finding—"the most damaging medical hoax of the last 100 years" (Flaherty, 2011)—had produced an "anti-vax" movement and declining vaccination rates. Instead of following the typical path toward disease elimination, U.S. measles rates in 2019 rose to their highest levels in 25 years (CDC, 2019; Graham et al., 2019). Unvaccinated children may suffer long-term harm or even death, as well as placing at risk those children too young to be fully vaccinated. Though the science self-corrected, the damage lingers on. Nevertheless, the good news is that scientific scrutiny, complete with replication, can inform and protect us.

Values in Psychology

Values affect what we study, how we study it, and how we interpret results. Researchers' values influence choice of research topics. Should we study worker productivity or worker morale? Sex discrimination or gender differences? Conformity or independence? Values can also color "the facts"—our observations and interpretations. Sometimes we see what we want or expect to see (**FIGURE 1.7**).

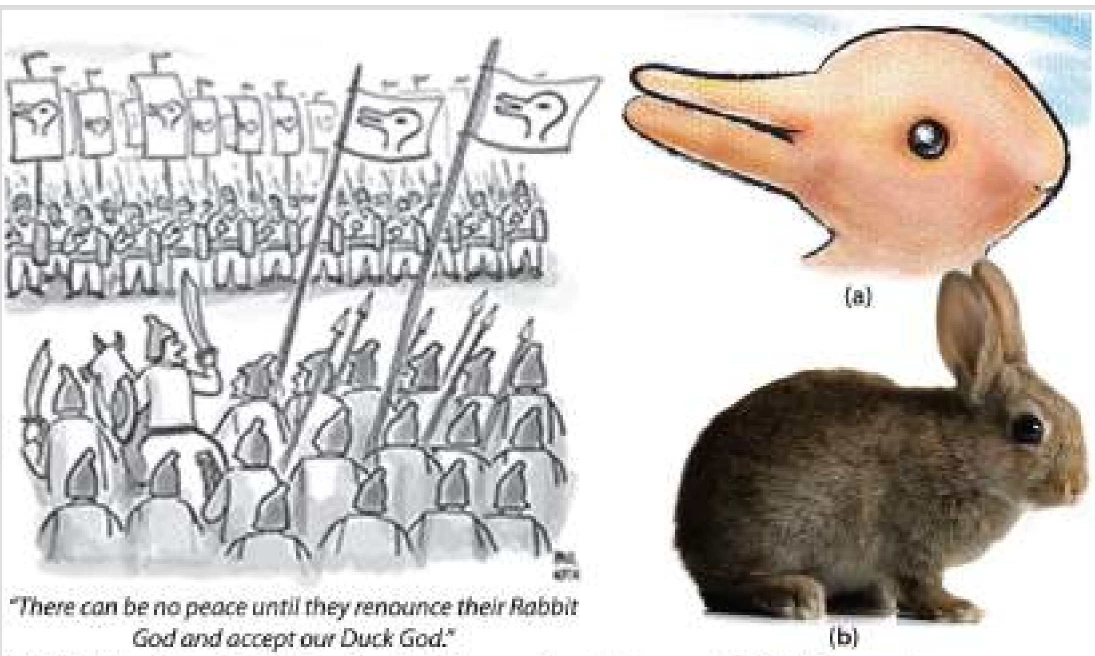


FIGURE 1.7 **What do you see?** Our expectations influence what we perceive in (a). Did you see a duck or a rabbit? Show some friends this image with the rabbit photo (b) covered up and see if they are more likely to perceive a duck. (Inspired by Shepard, 1990.)

Even the words we use to describe traits and tendencies can reflect our values. In psychology and in everyday speech, labels describe and labels evaluate: One person's *rigidity* is another's *consistency.* One person's *faith* is another's *fanaticism.* One person's *adultery* is another person's *open marriage.* Our labeling someone as *firm* or *stubborn, careful* or *picky, discreet* or *secretive* reveals our own attitudes.

So, values inform psychological science—and psychological science has the power to persuade. This may lead some to feel distrustful: Is psychology dangerously powerful? Might it be used to manipulate people? Knowledge, like all power, can be used for good or evil. Nuclear power has been used to light up cities—and to demolish them. Persuasive power has been used to educate people—and to deceive them. Although psychology does have the power to deceive, its purpose is to enlighten. Every day, psychologists explore ways to enhance learning, creativity, and compassion. Psychology speaks to many of our world's great problems—war, overpopulation, inequality, climate change, prejudice, family crises, crime—all of which involve attitudes and behaviors. Psychology also speaks to our deepest longings—for nourishment, for love, for happiness. Psychology cannot address all of life's great questions, but it speaks to some mighty important ones.

Left and middle: Macmillan Learning; right: Gordon Parks Foundation

Psychology speaks In making its historic 1954 school desegregation decision, the U.S. Supreme Court cited the expert testimony and research of psychologists Kenneth Clark and Mamie Phipps Clark (1947). The Clarks reported that, when given a choice between Black and White dolls, most African-American children chose the White doll, which indicated that they had likely absorbed and internalized anti-Black prejudice.

ASK YOURSELF

What other questions or concerns do you have about psychology?

RETRIEVAL PRACTICE

RP-14 How are animal and human research participants protected?

ANSWERS IN APPENDIX E

REVIEW Research Strategies: How Psychologists Ask and Answer Questions

LEARNING OBJECTIVES

Test Yourself Answer these repeated Learning Objective Questions on your own (before checking the answers in Appendix D) to improve your retention of the concepts (McDaniel et al., 2009, 2015).

LOQ 1-1: How does our everyday thinking sometimes lead us to a wrong conclusion?
LOQ 1-2: Why are we so vulnerable to believing untruths?
LOQ 1-3: How do theories advance psychological science?
LOQ 1-4: How do psychologists use case studies, naturalistic observations, and surveys to observe and describe behavior, and why is random sampling important?
LOQ 1-5: What does it mean when we say two things are correlated, and what are positive and negative correlations?
LOQ 1-6: What are *illusory correlations*, and what is *regression toward the mean?*
LOQ 1-7: Why do correlations enable prediction but not cause-effect explanation?
LOQ 1-8: What are the characteristics of experimentation that make it possible to isolate cause and effect?

LOQ 1-9: How would you know which research design to use?
LOQ 1-10: How can simplified laboratory conditions illuminate everyday life?
LOQ 1-11: Why do psychologists study animals, and what ethical research guidelines safeguard human and animal welfare? How do psychologists' values influence what they study and how they apply their results?

TERMS AND CONCEPTS TO REMEMBER

Test Yourself Write down the definition in your own words, then check your answer.

hindsight bias
theory
hypothesis
operational definition
replication
preregistration
meta-analysis
case study
naturalistic observation
survey
random sample
population
correlation
correlation coefficient
variable
scatterplot
illusory correlation
regression toward the mean
experiment
experimental group
control group
random assignment
double-blind procedure
placebo [pluh-SEE-bo] effect

independent variable
confounding variable
dependent variable
informed consent
debriefing

MASTER THE MATERIAL

Test Yourself Answer the following questions on your own first, then check your answers in Appendix E.

1. ________ ________ refers to our tendency to perceive events as obvious or inevitable after the fact.
2. As scientists, psychologists
   a. keep their methods private so others will not repeat their research.
   b. assume the truth of articles published in leading scientific journals.
   c. reject evidence that competes with traditional findings.
   d. are willing to ask questions and to reject claims that cannot be verified by research.
3. A theory-based prediction is called a(n) ________.
4. Which of the following is NOT one of the *descriptive* methods psychologists use to observe and describe behavior?
   a. A case study
   b. Naturalistic observation
   c. Correlational research
   d. A phone survey

5. For your survey, you need to establish a group of people who represent your country’s entire adult population. To do this, you will need to question a ________ sample of the population.
6. A study finds that the more childbirth training classes women attend, the less pain medication they require during childbirth. This finding can be stated as a ________ (positive/negative) correlation.
7. A ________ provides a visual representation of the direction and the strength of a relationship between two variables.
8. In a ________ correlation, the scores rise and fall together; in a(n) ________ correlation, one score falls as the other rises.
   a. positive; negative
   b. positive; illusory
   c. negative; weak
   d. strong; weak
9. How can regression toward the mean influence our interpretation of events?
10. Knowing that two events are correlated provides
    a. a basis for prediction.
    b. an explanation of why the events are related.
    c. proof that as one increases, the other also increases.
    d. an indication that an underlying third factor is at work.
11. Here are some recently reported correlations, with interpretations drawn by journalists. Knowing just these correlations, can you come up with other possible explanations for each of these?
    a. Alcohol use is associated with violence. (One interpretation: Drinking

triggers or unleashes aggressive behavior.)

b. Educated people live longer, on average, than less-educated people. (One interpretation: Education lengthens life and enhances health.)

c. Teens engaged in team sports are less likely to use drugs, smoke, have sex, carry weapons, and eat junk food than are teens who do not engage in team sports. (One interpretation: Team sports encourage healthy living.)

d. Adolescents who frequently see smoking in movies are more likely to smoke. (One interpretation: Movie stars' behavior influences impressionable teens.)

12. To explain behaviors and clarify cause and effect, psychologists use ______________.

13. To test the effect of a new drug on depression, we randomly assign people to control and experimental groups. Those in the control group take a pill that contains no medication. This pill is a ______________.

14. In a double-blind procedure,

    a. only the participants know whether they are in the control group or the experimental group.

    b. experimental and control group members will be carefully matched for age, sex, income, and education level.

    c. neither the participants nor the researchers know who is in the experimental group or control group.

    d. someone separate from the researcher will ask people to volunteer for the experimental group or the control group.

15. A researcher wants to determine whether noise level affects workers' blood pressure. In one group, she varies the level of noise in the environment and records participants' blood pressure. In this experiment, the level of noise is

the ____________ ____________.

16. The laboratory environment is designed to

    a. exactly re-create the events of everyday life.

    b. re-create psychological forces under *controlled* conditions.

    c. re-create psychological forces under *random* conditions.

    d. minimize the use of animals and humans in psychological research.

17. In defending their experimental research with animals, psychologists have noted that

    a. animals' physiology and behavior can tell us much about our own.

    b. animal experimentation sometimes helps animals as well as humans.

    c. animals are fascinating creatures and worthy of study.

    d. all of these statements are correct.

Continue testing yourself with **LearningCurve** or **Achieve Read & Practice** to learn and remember most effectively.

Statistical Reasoning in Everyday Life

In descriptive, correlational, and experimental research, statistics are tools that help us see and interpret what the unaided eye might miss. But accurate statistical understanding benefits everyone. To be an educated person today is to be able to apply simple statistical principles to everyday reasoning. One needn't memorize complicated formulas to think more clearly and critically about data.

Off-the-top-of-the-head estimates often misread reality and mislead the public. Someone throws out a big, round number. Others echo it, and before long the big, round number becomes public misinformation. Three examples:

- ***Ten percent of people are gay***. Or is it 2 to 4 percent, as suggested by various national surveys (Chapter 11)?
- ***We ordinarily use only 1 percent of our brain***. Or is it closer to 100 percent (Chapter 2)?
- ***To be healthy, walk 1 , steps a day.*** Or will 8,500 or 13,000 steps do the trick, or swimming or jogging (Mull, 2019)?

If you see an attention-grabbing headline presented without evidence—that nationally there are one million teen pregnancies, two million homeless seniors, or three million alcohol-related car accidents—you can be pretty sure that someone is guessing. If they want to emphasize the problem, they will be motivated to guess big. If they want to minimize the problem, they will guess small. *The point to remember* Use critical thinking when presented with big, round, undocumented numbers.

When setting goals, we love big, round numbers. We're far more likely to want to lose 20 pounds than 19 or 21 pounds (or an even 10 kilograms rather than 9.07 kilograms). And batters try to improve their batting average shortly before the season's end, making them nearly four times more likely to finish with a .300 average than with a .299 average (Pope & Simonsohn, 2011).

Statistical illiteracy also feeds needless health scares (Gigerenzer, 2010). In the 1990s, the British press reported a study showing that women taking a particular contraceptive pill had a 100 percent increased risk of blood clots that could produce strokes. The story went viral, causing thousands of women to stop taking the pill. What resulted? A wave of unwanted pregnancies and an estimated 13,000 additional abortions (which also are associated with increased blood-clot risk). Distracted by big, round numbers, few people focused on the study's actual findings: A 100 percent increased risk, indeed—but only from 1 in 7000 to 2 in 7000. Such false alarms underscore the need to think critically, to teach statistical reasoning, and to present statistical information more transparently.

Describing Data

LOQ 1-12

How do we describe data using three measures of central tendency, and what is the relative usefulness of the two measures of variation?

Once researchers have gathered their data, they may organize that data using *descriptive statistics.* One way to do this is to convert the data into a simple *bar graph,* as in **FIGURE 1.8**, which displays a distribution of different brands of trucks still on the road after a decade. When reading statistical graphs such as this one, take care. It's easy to design a graph to make a difference look big (Figure 1.8a) or small (Figure 1.8b). The secret lies in how you label the vertical scale (the *y-axis*).

The point to remember Think smart. When interpreting graphs, consider the scale labels and note their *range.*

ASK YOURSELF

Think of a time when you used statistics to make a point—maybe in class, in a paper, or in a discussion with a friend or family member. Looking back, were the data you cited credible and accurate? How do you know?

RETRIEVAL PRACTICE

RP-1 A truck manufacturer offered Figure 1.8's graph (a)—with actual brand names included—to suggest the much greater durability of its trucks. What does graph (b) make clear about the varying durability, and how is this accomplished?

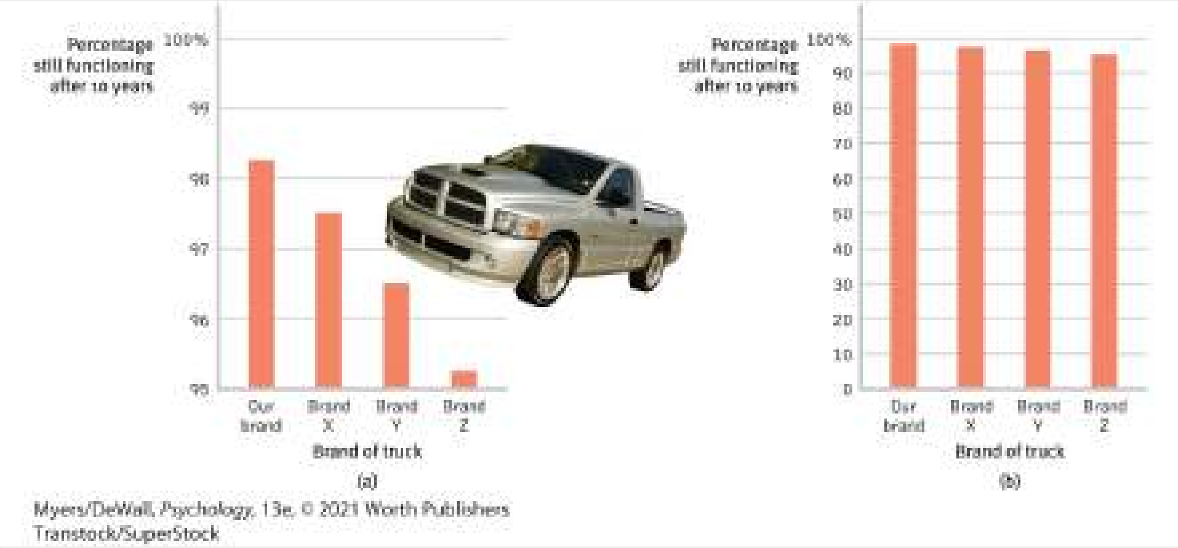


FIGURE 1.8 Read the scale labels

ANSWERS IN APPENDIX E

Measures of Central Tendency

The next step is to summarize the data using a *measure of central tendency,* a single score that represents a whole set of scores. The simplest measure is the **mode,** the most frequently occurring score or scores. The most familiar is the **mean,** or arithmetic average—the total sum of all the scores divided by the number of scores. The midpoint—the 50th percentile—is the **median**. On a divided highway, the median is the middle. So, too, with data: If you arrange all the scores in order from the highest to the lowest, half will be above the median and half will be below it.

mode
the most frequently occurring score(s) in a distribution.
mean
the arithmetic average of a distribution, obtained by adding the scores and then dividing by the number of scores.
median
the middle score in a distribution; half the scores are above it and half are below it.

Measures of central tendency neatly summarize data. But consider what happens to the mean when a distribution is lopsided—when it's *skewed* by a few way-out scores. With income data, for example, the mode, median, and mean often tell very different stories (**FIGURE 1.9**). This happens because the mean is biased by a few extreme incomes. When Amazon founder Jeff Bezos sits down in a small café, its average (mean) customer instantly becomes a billionaire. But median customer wealth remains unchanged. Understanding this, you can see why, according to the 2010 U.S. Census, nearly 65 percent of U.S. households have "below average" income. The bottom half of earners receive much less than half of the total national income. So, most Americans make less than average (the mean). Mean and median tell different true stories.

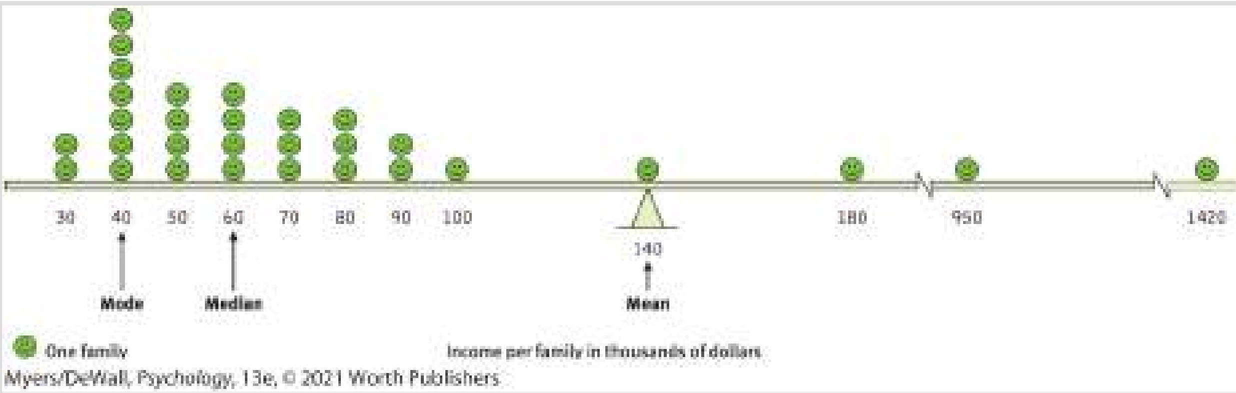


FIGURE 1.9 A skewed distribution This graphic representation of the distribution of a village's incomes illustrates the three measures of central tendency—mode, median, and mean. Note how just a few high incomes make the mean—the fulcrum point that balances the incomes above and below—deceptively high.

The point to remember Always note which measure of central tendency is reported. If it is a mean, consider whether a few atypical scores could be distorting it.

The average person has one ovary and one testicle.

Measures of Variation

Knowing the value of an appropriate measure of central tendency can tell us a great deal. But the single number omits other information. It helps to know something about the amount of *variation* in the data—how similar or diverse the scores are. Averages derived from scores with low variability are more reliable than averages based on scores with high variability. Consider a basketball player who scored between 13 and 17 points in each of the season's first 10 games. Knowing this, we would be more confident that she would score near 15 points in her next game than if her scores had varied from 5 to 25 points.

The **range** of scores—the gap between the lowest and highest—provides only a crude estimate of variation. A couple of extreme scores in an otherwise similar group, such as the $950,000 and $1,420,000 incomes in Figure 1.9, will create a deceptively large range.

range
the difference between the highest and lowest scores in a distribution.

The more useful standard for measuring how much scores deviate (differ) from one another is the **standard deviation.** It better gauges whether scores are packed together or dispersed, because it uses information from each score. The computation[4] assembles information about how much individual scores differ from the mean, which can be very telling. Let's say test scores from Class A and Class B both have the same mean (75 percent correct), but very different standard deviations (5.0 for Class A and 15.0 for Class B). Have you ever had test experiences like that—where two-thirds of your classmates in one course score in the 70 to 80 percent range, with scores in another course more spread out (two-thirds between 60 and 90 percent)? The standard deviation, as well as the mean score, tell us about how each class is faring.

standard deviation
a computed measure of how much scores vary around the mean score.

You can grasp the meaning of the standard deviation if you consider how scores naturally tend to be distributed. Large numbers of data—heights, intelligence scores, life expectancy (though not incomes)—often form a symmetrical, *bell-shaped* distribution. Most cases fall near the mean, and fewer cases fall near either extreme. This bell-shaped distribution is so typical that we call the curve it forms the **normal curve**.

normal curve
a symmetrical, bell-shaped curve that describes the distribution of many types of data; most scores fall near the mean (about 68 percent fall within one standard deviation of it) and fewer and fewer near the extremes. (Also called a *normal distribution.*)

As **FIGURE 1.10** shows, a useful property of the normal curve is that roughly 68 percent of the cases fall within one standard deviation on either side of the mean. About 95 percent of cases fall within two standard deviations. Thus, as Chapter 10 notes, about 68 percent of people taking an intelligence test will score within ±15 points of 100. About 95 percent will score within ±30 points.

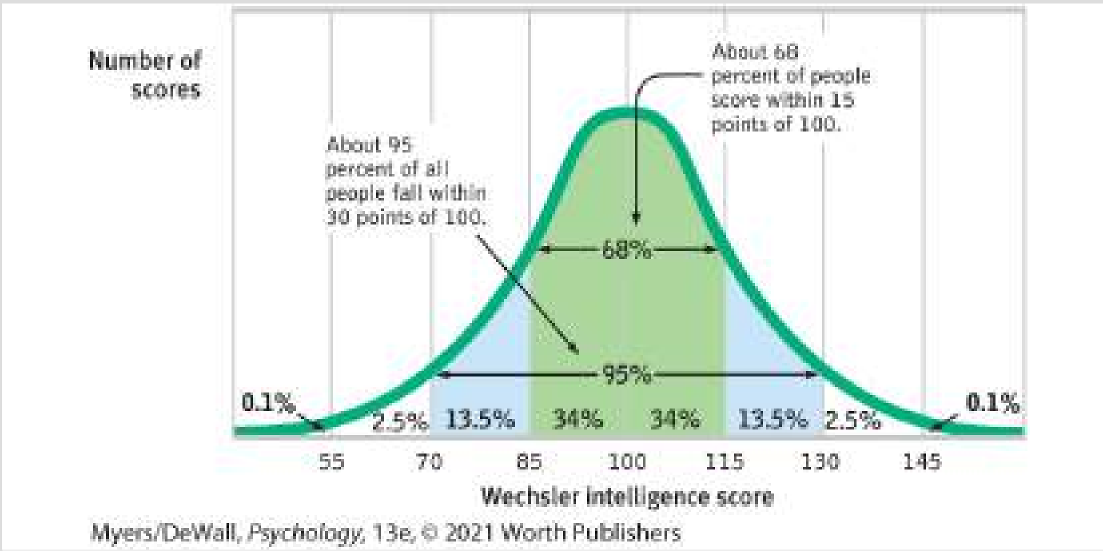


FIGURE 1.10 The normal curve Scores on aptitude tests tend to form a normal, or bell-shaped, curve. The most commonly used intelligence test, the Wechsler Adult Intelligence Scale, calls the average score 100.

RETRIEVAL PRACTICE

RP-2 The average of a distribution of scores is the ________. The score that shows up most often is the ________. The score right in the middle of a distribution (half the scores above it; half below) is the ________. We determine how much scores vary around the average in a way that includes information about the ________ of scores (difference between highest and lowest) by using the ________ ________ formula.

ANSWERS IN ***APPENDIX E***

For an interactive review of these statistical concepts, visit ***Topic Tutorial: PsychSim6, Descriptive Statistics.***

Significant Differences

LOQ 1-13

How do we know whether an observed difference can be generalized to other populations?

"The poor are getting poorer, but with the rich getting richer it all averages out in the long run."

The New Yorker Collection, 1988, Mirachi from cartoonbank.com. All Rights Reserved.

Data are "noisy." The average score in one group (those who deactivated their Facebook account, in the experiment we mentioned earlier) could conceivably differ from the average score in another group (those who didn't) not because of any real difference, but merely because of chance fluctuations in the people sampled. How confidently, then, can we *infer* that an observed difference is not just a fluke—a chance result from the research sample? For guidance, we can ask how reliable and statistically significant the differences are. These *inferential statistics* help us determine if results can be generalized to a larger population (all those in a group being studied).

When Is an Observed Difference Reliable?

In deciding when it is safe to generalize from a sample, we should keep three principles in mind:

1. ***Representative samples are better than biased (unrepresentative) samples.*** The best basis for generalizing is not from the exceptional and memorable cases one finds at the extremes but from a representative sample of cases. Research never randomly samples the whole human population. Thus, it pays to keep in mind what population a study has sampled.
2. ***Less-variable observations are more reliable than those that are more variable.*** As we noted earlier in the example of the basketball player whose game-to-game points were consistent, an average is more reliable when it comes from scores with low variability.
3. ***More cases are better than fewer.*** An eager prospective student visits two universities, each for a day. At the first, the student randomly attends two classes and finds both instructors to be witty and engaging. At the second, the two sampled instructors seem dull and uninspiring. Returning home, the student (discounting the small sample size of only two teachers at each institution) tells friends about the "great teachers" at the first school and the "bores" at the second. Again, we know it but we ignore it: *Averages based on many cases are more reliable* (less variable) than averages based on only a few cases. After noticing that small schools were overrepresented among the most successful schools, several foundations invested in splitting larger schools into smaller ones—without realizing that small schools were also overrepresented among the *least* successful, because schools with fewer students have more variable outcomes (Nisbett, 2015). Again, more cases make for a more reliable average and a more replicable study.

The point to remember Smart thinkers are not overly impressed by a few anecdotes. Generalizations based on a few unrepresentative cases are unreliable.

When Is an Observed Difference Significant?

Let's say you compared men's and women's scores on a laboratory test of aggression. You found that men behaved more aggressively than women. But individuals differ. How likely is it that your observed gender difference was just a fluke?

Researchers use statistical steps to answer this question. Statistical tests begin with the assumption that no differences exist between groups, an assumption called the *null* *hypothesis.* Using statistics we may conclude that the gender difference we observed is so large that it's unlikely to fit the null hypothesis. That's when we reject the null hypothesis of no differences, and we say that the result is **statistically significant**. Such a large difference would support an *alternative hypothesis*—that a difference (in aggression, for our example) does exist between the groups being studied (men and women).

statistical significance
a statement of how likely it is that a result (such as a difference between samples) occurred by chance, assuming there is no difference between the populations being studied.

How does the size of the difference between them—the *effect size*—determine statistical significance? First, when averages from two samples are each reliable measures of their respective populations (as when each is based on many observations that have low variability), then any difference between the two samples is more likely to be statistically significant. (For our example: The less the variability in women's and in men's aggression scores, the more confidence we would have that our observed gender difference is real.) When the difference between the sample averages is *large* (as long as the samples are based on many observations), we also have more confidence that the difference between them reflects a real difference in their populations.

In short, when samples are large, and when the difference between them is relatively large, we say the difference has statistical significance. This means that the observed difference is probably more than just chance variation between the samples, and we can reject the null hypothesis of no existing differences.

In judging statistical significance, psychologists are conservative. They are like juries that must presume innocence until guilt is proven. Many psychological tests provide *p*-values, which indicate the probability of the null hypothesis given the sample data. For most psychologists, proof beyond a reasonable doubt means not making much of a finding unless the probability (*p*-value) of the null hypothesis is less than 5 percent ($p < .05$). Some researchers argue that statistical significance is overemphasized, noting that a "nonsignificant" result does *not* mean—as people often assume—a complete lack of difference between the groups (Amrhein et al., 2019). It only indicates greater uncertainty. For now, many psychologists continue to use $p < .05$, but stay tuned.

When learning about research, you should remember that, given large enough or homogeneous enough samples, a difference between them may be “statistically significant” yet have little *practical* significance. They might be statistically significant but have a small effect size. Comparisons of intelligence test scores among hundreds of thousands of first-born and later-born individuals indicate a highly significant tendency for first-born individuals to have higher average scores than their later-born siblings (Rohrer et al., 2015; Zajonc & Markus, 1975). But because the scores differ only slightly, the “significant” difference has a small effect size and little practical importance.

Reprinted by permission of Andres McMeel Syndicate, Inc.

The point to remember Statistical significance indicates the *likelihood* that a result would happen by chance if the null hypothesis (of no difference) were true. But this does not say anything about the *importance* of the result.

ASK YOURSELF

Can you think of a situation where you were fooled by a writer or speaker’s attempts to persuade you with statistics? What have you learned in this chapter that will be most helpful in the future to avoid being misled?

RETRIEVAL PRACTICE

RP-3 Can you solve this puzzle?

The registrar’s office at the University of Michigan has found that about 100 students in Arts and Sciences usually have perfect marks at the end of their first term. However, only about 10 to 15 students graduate with perfect marks. What do you think is the most likely explanation for the fact that there are more perfect marks after one term than at graduation (Jepson et al., 1983)?

RP-4 ________ statistics summarize data, while ________ statistics determine if data can be generalized to other populations.

ANSWERS IN APPENDIX E

For a review, engage online with ***Concept Practice: Statistical Significance.***

REVIEW Statistical Reasoning in Everyday Life

LEARNING OBJECTIVES

Test Yourself Answer these repeated Learning Objective Questions on your own (before checking the answers in Appendix D) to improve your retention of the concepts (McDaniel et al., 2009, 2015).

LOQ 1-12: How do we describe data using three measures of central tendency, and what is the relative usefulness of the two measures of variation?
LOQ 1-13: How do we know whether an observed difference can be generalized to other populations?

TERMS AND CONCEPTS TO REMEMBER

Test Yourself Write down the definition in your own words, then check your answer.

mode
mean
median

range
standard deviation
normal curve
statistical significance

MASTER THE MATERIAL

Test Yourself Answer the following questions on your own first, then check your answers in Appendix E.

1. Which of the three measures of central tendency is most easily distorted by a few very high or very low scores?
   a. The mode
   b. The mean
   c. The median
   d. They are all equally vulnerable to distortion from atypical scores.
2. The standard deviation is the most useful measure of variation in a set of data because it tells us
   a. the difference between the highest and lowest scores in the set.
   b. the extent to which the sample being used deviates from the bigger population it represents.
   c. how much individual scores differ from the mode.
   d. how much individual scores differ from the mean.
3. Another name for a bell-shaped distribution, in which most scores fall near the middle and fewer scores fall at each extreme, is a ____________ ____________.

4. When sample averages are ____________ and the difference between them is ____________, we can say the difference is more likely to be statistically significant.

   a. reliable; large

   b. reliable; small

   c. due to chance; large

   d. due to chance; small

Continue testing yourself with **LearningCurve** or **Achieve Read & Practice** to learn and remember most effectively.